Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem

Charles H. Bennett*, Peter W. Shor†,
John A. Smolin*, and Ashish V. Thapliyal‡

Abstract

We show that the entanglement-assisted classical capacity (C_E) of a noisy quantum channel is given by an expression parallel to that for the capacity of a purely classical channel: i.e., the maximum, over channel inputs ρ, of the entropy of the channel input plus the entropy of the channel output minus their joint entropy, the latter being defined as the entropy of an entangled purification of ρ after half of it has passed through the channel. We calculate entanglement-assisted capacities for the amplitude damping channel, and for bosonic channels in the presence of attenuation and Gaussian noise. We discuss how many independent parameters are required to completely characterize the asymptotic behavior of a general quantum channel, alone or in the presence of ancillary resources such as prior entanglement. In the classical analog of entanglement assisted communication—communication over a discrete memoryless channel (DMC) between parties who share prior random information—we show that one parameter is sufficient, i.e., that in the presence of prior shared random information, all DMC's of equal capacity can simulate one another with unit asymptotic efficiency.

* IBM T.J. Watson Reseach Center, Yorktown Heights, NY 10598, USA, †AT&T Labs – Research, Florham Park, NJ 07932, USA, ‡ Dept. of Physics, U.C. Santa Barbara, Santa Barbara, CA 93106, USA

1 Introduction

A (memoryless) quantum communications channel can be viewed physically as a process wherein a quantum system interacts with an environment (which may be taken to initially be in a standard state) on its way from a sender to a receiver; it may be defined mathematically as a completely positive, trace-preserving linear map on density operators. The theory of quantum channels is richer and less well understood than that of classical channels. For example, quantum channels have several distinct capacities, depending on what one is trying to use them for, and what additional resources are brought into play. These include

- The ordinary classical capacity C, defined as the maximum asymptotic rate at which classical bits can be transmitted reliably through the channel, with the help of a quantum encoder and decoder.
- The ordinary quantum capacity Q, which is the maximum asymptotic rate at which qubits can be transmitted under similar circumstances.
- The classically assisted quantum capacity Q_2, which is the maximum asymptotic rate of reliable qubit transmission with the help of unlimited use of a 2-way classical side channel between sender and receiver.
- The entanglement assisted classical capacity C_E, which is the maximum asymptotic rate of reliable bit transmission with the help of unlimited prior entanglement between the sender and receiver.

Somewhat unexpectedly, the last of these has turned out to be the simplest to calculate, because, as we show in section II, it is given by an expression analogous to the formula expressing the classical capacity of a classical channel as the maximum, over input distributions, of the input:output mutual information. Section III calculates entanglement assisted capacities of the amplitude damping channel and of Gaussian noise/attenuating bosonic channels.

We return now to a general discussion of quantum channels and capacities, in order to provide motivation for section IV of the paper, on what we call the reverse Shannon theorem.

Aside from the constraints $Q \le C \le C_E$, and $Q \le Q_2$, which are obvious consequences of the definitions, the four capacities appear to vary rather independently. It is conjectured that $Q_2 \le C$, but this has not been proved to date. Except in special cases, it is not possible, without knowing the parameters of a channel, to infer any one of its four capacities from the other three.

Channel	Q	Q_2	C	C_E
Noiseless qubit channel	1	1	1	2
50% erasure qubit channel	0	1/2	1/2	1
2/3 depolarizing qubit channel	0	0	0.0817	0.2075
Noiseless bit channel = 100% dephasing qubit channel	0	0	1	1

Table 1: Capacities of several quantum channels. The value 0.0817 for the 2/3-depolarizing qubit channel is the channel's one-shot capacity, which would be equal to C if, as generally believed, classical capacity is additive for this channel.

This independence is illustrated in Table 1, which compares the capacities of several simple channels for which they are known exactly. The channels incidentally illustrate four different degrees of qualitative quantumness: the first can carry qubits unassisted, the second requires classical assistance to do so, the third has no quantum capacity at all but still exhibits quantum behavior in that its capacity is increased by entanglement, while the fourth is completely classical, and so unaffected by entanglement. Contrary to an earlier conjecture of ours, we have found channels for which $Q > 0$ but $C = C_E$. One example is a channel mapping three qubits to two qubits which is switched between two different behaviors by the first input qubit. The channel operates as follows: The first qubit is measured in the $|0\rangle$, $|1\rangle$ basis. If the result is $|0\rangle$, then the other two qubits are dephased and transmitted as classical bits; if the result is $|1\rangle$, the first qubit is transmitted intact and the second qubit is replaced by the completely mixed state. This channel has $Q = Q_2 = 1$ and $C = C_E = 2$.

This complex situation naturally raises the question of how many independent parameters are needed to characterize the important asymptotic, capacity-like properties of a general quantum channel. A full understanding of quantum channels would enable us to calculate not only their capacities, but more generally, for any two channels $\mathcal{M}$ and $\mathcal{N}$, the asymptotic efficiency (possibly zero) with which $\mathcal{M}$ can simulate $\mathcal{N}$, both alone and in the presence of ancillary resources such as classical communication or shared entanglement.

One motivation for studying communication in the presence of ancillary resources is that it can simplify the classification of channels' capacities to simulate one another. This is so because if a simulation is possible without the ancillary resource, then the simulation remains possible with it, though not necessarily vice versa. For example, Q and C represent a channel's asymptotic efficiencies of simulating, respectively, a noiseless qubit channel and a noiseless classical bit channel. In the absence of ancillary resources these two capacities can vary independently, subject to the constraint $Q \leq C$, but in the presence of unlimited prior entanglement, the relation between them becomes fixed: $C_E = 2Q_E$, because entanglement allows a noiseless 2-bit classical channel to simulate a noiseless 1-qubit channel and vice versa (via teleportation and superdense coding).

We conjecture that prior entanglement so simplifies the complex landscape of quantum channels that only a single free parameter remains. This would mean that, in the presence of unlimited prior entanglement, any two quantum channels of equal C_E could simulate one another with unit asymptotic efficiency. Section IV proves a classical analog of this conjecture, namely that in the presence of prior random information shared between sender and receiver, any two discrete memoryless classical channels (DMC's) can simulate one another with unit asymptotic efficiency. We call this the classical reverse Shannon theorem because it establishes the ability of a noiseless classical DMC's to simulate noisy ones of equal capacity, whereas the ordinary Shannon theorem establishes that noisy DMC's can simulate noiseless ones of equal capacity.

Another ancillary resource—classical communication—also simplifies the landscape of quantum channels, but probably not so much. The presence of unlimited classical communication does allow certain otherwise inequivalent pairs of channels to simulate one another (for example, a noiseless qubit channel and a 50% erasure channel on 4-dimensional Hilbert space), but it does not render all channels of equal Q_2 asymptotically equivalent. So-called bound-entangled channels [15, 11] have $Q_2 = 0$, but unlike classical channels (which also have $Q_2 = 0$) they can be used to prepare bound entangled states. Because the distinction between bound entangled and unentangled states does not vanish asymptotically [23], bound-entangled and classical channels must be asymptotically inequivalent, despite having the same Q_2.

The various capacities of a quantum channel $\mathcal{N}$ may be defined within a common framework,

$$C_X(\mathcal{N}) = \lim_{\epsilon\rightarrow 0} \limsup_{n\rightarrow\infty} \{\frac{m}{n} : \exists_{\mathcal{A}} \exists_{\mathcal{B}} \forall_{\psi\in\Gamma_m} F(\psi, \mathcal{A}, \mathcal{B}, \mathcal{N}) \ > 1-\epsilon \ \}. \quad (1)$$

Here C_X is a generalized capacity; $\mathcal{A}$ is an encoding subprotocol, to be performed by Alice, which receives an m-qubit state ψ belonging to some set Γ_m of allowable inputs to the entire protocol, and produces n possibly entangled inputs to the channel $\mathcal{N}$; $\mathcal{B}$ is a decoding subprotocol, to be performed by

Bob, which receives n (possibly entangled) channel outputs and produces an m-qubit output for the entire protocol; finally $F(\psi, \mathcal{A}, \mathcal{B}, \mathcal{N})$ is the *fidelity* of this output relative to the input ψ, i.e., the probability that the output state would pass a test determining whether it is equal to the input. Different capacities are defined depending on the specification of Γ, $\mathcal{A}$ and $\mathcal{B}$. The classical capacities C and C_E are defined by restricting ψ to a standard orthonormal set of states, without loss of generality the "Boolean" states labeled by bit strings $\Gamma_m = \{|0\rangle, |1\rangle\}^{\otimes m}$; for the quantum capacities Q and Q_2, Γ_m is the entire 2^m dimensional Hilbert space $\mathcal{H}_2^{\otimes m}$. For the simple capacities Q and C, the Alice and Bob subprotocols are completely-positive trace-preserving maps from $\mathcal{H}_2^{\otimes m}$ to the input space of $\mathcal{N}^{\otimes n}$, and from the output space of $\mathcal{N}^{\otimes n}$ back to $\mathcal{H}_2^{\otimes m}$. For C_E and Q_2, the subprotocols are more complicated, in the first case drawing on a supply of ebits (maximally entangled pairs of qubits) shared beforehand between Alice and Bob, and in the latter case making use of a 2-way classical channel between Alice and Bob. The definition of Q_2 thus includes interactive protocols, in which the n channel uses do not take place all at once, but may be interspersed with rounds of classical communication.

The classical capacity of a classical discrete memoryless channel is also given by an expression of the same form, with ψ restricted to Boolean values; the encoder $\mathcal{A}$, decoder $\mathcal{B}$, and channel $\mathcal{N}$ all being restricted to be classical stochastic maps; and the fidelity F being defined as the probability that the (Boolean) output of $\mathcal{B}(\mathcal{N}^{\otimes n}(\mathcal{A}(\psi)))$ is equal to the input ψ. We will sometimes indicate these restrictions implicitly by using upper case italic letters (e.g. N) for classical stochastic maps, and lower case italic letters (e.g. x) for classical discrete data. The definition of classical capacity would then be

$$C(N) = \lim_{\epsilon\rightarrow 0} \limsup_{n\rightarrow\infty} \{\frac{m}{n} : \exists_A \exists_B \ \forall_{x\in\{0,1\}^m} F(x, A, B, N) \ > 1-\epsilon \ \}. \qquad (2)$$

A classical stochastic map, or classical channel, may be defined in quantum terms as one that is completely dephasing in the Boolean basis both with regard to its inputs and its outputs. A channel, in other words, is classical if and only if it can be represented as a composition $\mathcal{D}'\mathcal{G}\mathcal{D}$ of a completely dephasing channel $\mathcal{D}$ on the input Hilbert space, followed by a general quantum channel $\mathcal{G}$, followed by a completely dephasing channel $\mathcal{D}'$ on the output Hilbert space (a dephasing channel is one that makes a von Neumann measurement in the Boolean basis and resends the result of the measurement). Dephasing only the inputs, or only the outputs, is in general insufficient to abolish all quantum properties of a quantum channel $\mathcal{G}$.

The notion of capacity may be further generalized to define a capacity of one channel $\mathcal{N}$ to simulate another channel $\mathcal{M}$. This may be defined as

$$C_X(\mathcal{N},\mathcal{M}) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \{ \frac{m}{n} : \exists_{\mathcal{A},\mathcal{B}} \forall_{\psi\in H_M^{\otimes m}} F(\mathcal{M}^{\otimes m}(\psi),\mathcal{A},\mathcal{B},\mathcal{N}) \ > 1-\epsilon \ \}, \tag{3}$$

where $\mathcal{A}$ and $\mathcal{B}$ are respectively Alice's and Bob's subprotocols which together enable Alice to receive an input ψ in $H_M^{\otimes m}$ (the tensor product of m copies of the input Hilbert space H_M of the channel $\mathcal{M}$ to be simulated) and, making n forward uses of the simulating channel $\mathcal{N}$, allow Bob to produce some output state, and $F(\mathcal{M}^{\otimes m}(\psi),\mathcal{A},\mathcal{B},\mathcal{N})$ is the fidelity of this output state with respect to the state that would have been generated by sending the input ψ through $\mathcal{M}^{\otimes m}$.

These definitions of capacity are all asymptotic, depending on the properties of $\mathcal{N}^{\otimes n}$ in the limit $n\to\infty$. However, several of the capacities are given by, or closely related to, non-asymptotic expressions involving input and output entropies for a single use of the channel. Figure 1 shows a scenario in which a quantum system Q, initially in mixed state ρ, is sent through the channel, emerging in a mixed state $\mathcal{N}(\rho)$. It is useful to think of the initial mixed state as being part of an entangled pure state Φ_ρ^{QR} where R is some reference system that is never operated upon physically. Similarly the channel can be thought of as a unitary interaction U between the quantum system Q and some environment subsystem E, which is initially supplied in a standard pure state 0^E, and leaves the interaction in a mixed state $\mathcal{E}(\rho)^E$. Thus $\mathcal{N}$ and $\mathcal{E}$ are completely positive maps relating the final states of the channel output and environment, respectively, to the initial state of the channel input, when the initial state of the environment is held fixed. The mnemonic superscripts Q, R, E indicate, when necessary, to what system a density operator refers.

Under these circumstances three useful von Neumann entropies may be defined, the input entropy

$$H(\rho^Q) = -\mathrm{tr}\rho^Q \log_2 \rho^Q,$$

the output entropy

$$H(\mathcal{N}(\rho)^Q),$$

and the *entropy exchange*

$$H((\mathcal{N}\otimes\mathcal{I})\Phi_\rho^{QR}) = H(\mathcal{E}(\rho)^E).$$

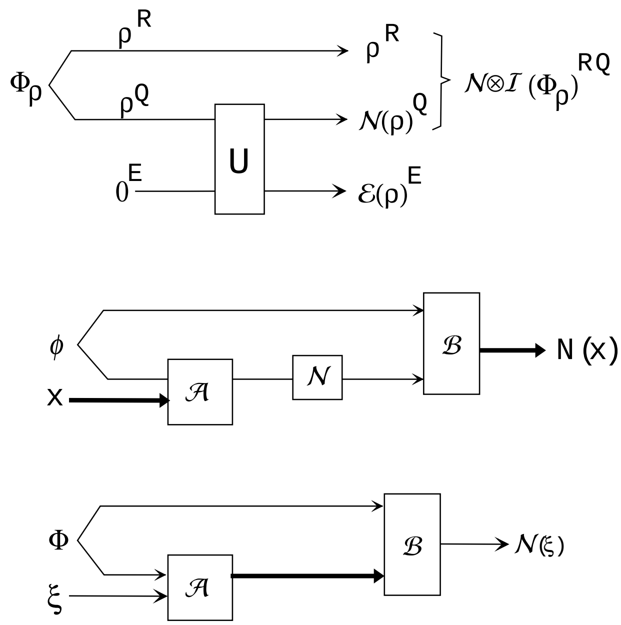

Figure 1: Top: A quantum system Q in mixed state ρ is sent through the noisy channel $\mathcal{N}$, which may be viewed as a unitary interaction U with an environment E. Meanwhile a purifying reference system R is sent through the identity channel $\mathcal{I}$. The final joint state of RQ has the same entropy as the final state $\mathcal{E}(\rho)$ of the environment. Middle: Entanglement assisted classical communication. Half of an entangled state ϕ is acted on by classical information x (thick line), then sent through the noisy channel $\mathcal{N}$, after which it is decoded, in conjunction with the other half of ϕ, to yield a classical output $\mathtt{N}(x)$. The entanglement-assisted classical capacity of $\mathcal{N}$ is defined as the maximum, over entangled states ϕ and quantum operations $\mathcal{A}$, of the capacity of the classical channel N. Bottom: Simulation of a noisy quantum channel $\mathcal{N}$ by noiseless classical communication and prior shared entanglement. The classical simulation cost of $\mathcal{N}$ is the minimum number of bits of forward classical communication required to simulate $\mathcal{N}$ in the presence of prior entanglement.

The complicated left side of the last equation represents the entropy of the joint state of the subsystem Q which has been through the channel, and the reference system R, which has not, but may still be more or less entangled with it. The density operator $(\mathcal{N} \otimes \mathcal{I})\Phi_\rho$ is the quantum analog of a joint input:output probability distribution, because it has $\mathcal{N}(\rho)$ and ρ as its partial traces. Without the reference system, the notion of a joint input:output mixed state would be problematic, because the input and output are not present at the same time, and the no-cloning theorem prevents Alice from retaining a spare copy of the input to be compared with the one sent through the channel. The entropy exchange is also equal to the final entropy of the environment $H(\mathcal{E}(\rho))$, because the tripartite system QRE remains throughout in a pure state; making its two complementary subsystems E and QR always isospectral. The relations between these entropies and quantum channels have been well reviewed by Schumacher and Nielsen [21] and by Holveo and Werner [13].

By Shannon's theorem, the capacity of a classical channel N is the maximum, over input distributions, of the input:output mutual information, in other words of the input entropy, plus the output entropy, minus the joint entropy of input and output. The quantum generalization of mutual information for a bipartite mixed state ρ^{AB}, which reduces to classical mutual information when ρ^{AB} is diagonal in a product basis of the two subsystems, is

$$H(\rho^A) + H(\rho^B) - H(\rho^{AB}).$$

where

$$\rho^A = \mathrm{tr}_B \rho^{AB} \qquad \text{and} \qquad \rho^B = \mathrm{tr}_A \rho^{AB}$$

Thus, in terms of Fig, 1, the classical capacity of a classical channel can be expressed as

$$C(N) = \max_{\rho \in \Delta} H(\rho) + H(N(\rho)) - H((N \otimes \mathcal{I})(\Phi_\rho)) \tag{4}$$

where Δ is the class of density operators on the channel's input Hilbert space that are diagonal in the Boolean basis, and N, as a classical channel, preserves diagonality in this basis. The third term (entropy exchange), for a classical channel N, is just the joint Shannon entropy of the classically correlated Boolean input and output, because the von Neumann entropies reduce to Shannon entropies when evaluated in the Schmidt basis of Φ_ρ, with respect to which all states are diagonal. The restriction to classical inputs $\rho \in \Delta$ can be removed, because any non-diagonal elements in ρ would only reduce the first term, while leaving the other two terms unchanged, by virtue of the diagonality-enforcing properties of the channel.

Thus, the expression

$$\max_{\rho \in \mathcal{H}_{\text{in}}} H(\rho) + H(\mathcal{N}(\rho)) - H((\mathcal{N} \otimes \mathcal{I})\Phi_\rho), \tag{5}$$

is a natural generalization to quantum channels $\mathcal{N}$ of a classical channel's maximal input:output mutual information, and it is equal to the classical capacity whenever $\mathcal{N}$ is classical, as defined previously in this section.

One might suppose that this expression continues to give the classical capacity of a general quantum channel $\mathcal{N}$, but that is not so, as can be seen by considering the simple case $\mathcal{N} = \mathcal{I}$ of a noiseless qubit channel. Here the maximum is attained on a uniform input mixed state $\rho = I/2$, causing the first two terms each to have the value 1, while the last term is zero, giving a total of 2. This is not the ordinary classical capacity of the noiseless qubit channel, but rather its entanglement-assisted capacity $C_E(\mathcal{N})$. In the next section we show that this is true of quantum channels in general, as stated by the following theorem.

Theorem 1 *Given a quantum channel $\mathcal{N}$, then the entanglement-assisted capacity of the quantum channel C_E is equal to the quantum mutual information*

$$C_E = \max_{\rho \in \mathcal{H}_{\text{in}}} H(\rho) + H(\mathcal{N}(\rho)) - H((\mathcal{N} \otimes \mathcal{I})\Phi_\rho). \tag{6}$$

Here the capacity C_E is defined as the supremum of Eq. (1) when ψ ranges over Boolean states and $\mathcal{A}$, $\mathcal{B}$ over all protocols where Alice and Bob start with an arbitrarily large number of shared EPR pairs, but have no access to any communication channels other than $\mathcal{N}$.

Another capacity theorem which has been proven for quantum channels is the Holevo-Schumacher-Westmoreland theorem[14, 22], which says that if the signals that Bob receives are constrained to lie in a set of quantum states ρ_i', where Alice chooses i (for example, by supplying input state ρ_i to the channel $\mathcal{N}$) then the capacity is given by

$$C_H = \max_{p_i | \sum_i p_i = 1} \quad H(\sum_i p_i \rho_i) - \sum_i p_i H(\rho_i). \tag{7}$$

This gives a means to calculate a constrained classical capacity for a quantum channel $\mathcal{N}$ if the sender is not allowed to use entangled inputs: the channel's Holevo capacity $C_H(\mathcal{N})$ being defined as the maximum of $C_H(\{\mathcal{N}(\rho_i)\})$ over all possible sets of input states $\{\rho_i\}$. We will be using this theorem extensively in the proof of our entanglement-assisted capacity bound.

In our original paper [5], we proved the formula (5) for certain special cases, including the depolarizing channel and the erasure channel. We did this by sandwiching the entanglement-assisted capacity between two other capacities, which for certain channels turned out to be equal. The higher of these two capacities we called the forward classical communication cost via teleportation, ($FCCC_{Tp}$), which is the amount of forward classical communication needed to simulate the channel $\mathcal{N}$ by teleporting over a noisy classical channel. The lower of these two bounds we called C_{Sd}, which is the capacity obtained by using the noisy quantum channel $\mathcal{N}$ in the superdense coding protocol. We have that $C_{Sd} \le C_E \le FCCC_{Tp}$. Thus, if $C_{Sd} = FCCC_{Tp}$ for a channel, we have obtained the entanglement-assisted capacity of the channel. In order for this argument to work, we needed the classical reverse Shannon theorem, which says that a noisy classical channel can be simulated by a noiseless classical channel of the same capacity, as long as the sender and receiver have access to shared random bits. We needed this theorem because the causality argument showing that EPR pairs do not add to the capacity of a classical channel appears to work only for noiseless channels. We sketched the proof of the classical reverse Shannon theorem in our previous paper, and give it in full in this paper.

In our previous paper, the bounds C_{Sd} and $FCCC_{Tp}$ are both computed using single-symbol protocols; that is, both the superdense coding protocol and the simulation of the channel by teleportation via a noisy classical channel are carried out with a single use of the channel. The capacity is then obtained using the classical Shannon formula for a classical channel associated with these protocols. In this paper, we obtain bounds using multiple-symbol protocols, which perform entangled operations on many uses of the channel. We then perform the capacity computations using the Holevo-Schumacher-Westmoreland formula (7).

2 Formula for Entanglement Assisted Classical Capacity

Assume we have a quantum channel $\mathcal{N}$ which maps a Hilbert space $\mathcal{H}_{\text{in}}$ to another Hilbert space $\mathcal{H}_{\text{out}}$. Let C_E be the classical capacity of the channel when the sender and receiver have an unbounded supply of EPR pairs to use in the communication protocol. This section proves that the entanglement-assisted capacity of a channel is the maximum quantum mutual information attainable between the two parts of an entangled quantum state, one part of which has been passed through the channel. That is,

$$C_E(\mathcal{N}) = \max_{\rho \in \mathcal{H}_{\text{in}}} \quad H(\rho) + H(\mathcal{N}(\rho)) - H((\mathcal{N} \otimes \mathcal{I})\Phi_\rho), \tag{8}$$

where $H(\rho)$ denotes the von Neumann entropy of a density matrix $\rho \in \mathcal{H}_{\text{in}}$, $H(\mathcal{N}(\rho))$ denotes the von Neumann entropy of the output when ρ is input into the channel, and $H((\mathcal{N} \otimes \mathcal{I})\Phi_\rho)$ denotes the von Neumann entropy of a standard purification Φ_ρ of ρ, half of which has been sent through the channel $\mathcal{N}$ while the other half has been sent through the identity channel $\mathcal{I}$ (this corresponds to the portion of the entangled state that Bob holds at the start of the protocol). Here, we have $\Phi_\rho \in \mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{anc}}$ and $\text{Tr}_{\text{anc}} \Phi_\rho = \rho$. Since all purifications give the same entropy in this formula, we need not specify which one we use. As pointed out earlier, the right hand side of Eq. (8) parallels the expression for capacity of a classical channel as the maximum, over input distributions, of the input:output mutual information.

Lindblad [17], Barnum et al[2], and Adami and Cerf [8] characterized several important properties of the quantum mutual information, including positivity and additivity. Adami and Cerf argued that the right side of Eq. (8) represents an important channel property, calling it the channel's "von Neumann capacity", but they did not indicate what kind of communication task this capacity represented the channel's asymptotic efficiency for doing. Now we know that it is the channel's efficiency for transmitting classical information when the sender and receiver share prior entanglement.

In our demonstration that Eq. (8) is indeed the correct expression for entanglement assisted classical capacity, the first subsection gives an entanglement assisted classical communication protocol which can asymptotically achieve the rate $C_E(\mathcal{N}) - \epsilon$ for any ϵ. The second subsection gives a proof of a crucial lemma needed in the first subsection. The third subsection shows that the right hand side of Eq. (8) is indeed an upper bound for $C_E(\mathcal{N})$. The fourth subsection proves several entropy inequalities that were needed in the third subsection.

2.1 Proof of the Lower Bound

We first show the inequality

$$C_E(\mathcal{N}) \geq \max_{\rho \in \mathcal{H}_{\text{in}}} \quad H(\rho) + H(\mathcal{N}(\rho)) - H(\mathcal{N} \otimes \mathcal{I}\ (\Phi_\rho)). \tag{9}$$

for $\rho = \frac{1}{d} I$, where $d = \dim \mathcal{H}_{\text{in}}$ and I is the identity matrix. We do this by using Holevo's formula for quantum capacity. The protocol we use is the same as that given in our earlier paper on C_E, although our proof is different [5]. To prove the inequality (9), we need the generalization of the Pauli matrices to d dimensions. There are d^2 of these matrices, which are given by $U_{j,k} = T^j R^k$. Here the matrices T and R are defined by their entries

$$T_{a,b} = \delta_{a,\,b-1 \bmod d} \quad \text{and} \quad R_{a,b} = e^{2\pi i a/d}\delta_{a,b} \tag{10}$$

as in [1]. To achieve the capacity given by the above formula, Alice and Bob start by sharing a d-dimensional maximally entangled state ϕ. Alice applies one of the d^2 transformations $U_{j,k}$ to her part of ϕ, and then sends it through the channel $\mathcal{N}$. Bob gets one of the d^2 quantum states $(\mathcal{N}\otimes\mathcal{I})(U_{j,k}\otimes\mathcal{I})\phi$. It is straightforward to show that averaging over the matrices $U_{j,k}$ effectively disentangles Alice's and Bob's pieces, so we obtain

$$\sum_k (\mathcal{N}\otimes\mathcal{I})(U_{j,k}\otimes\mathcal{I})\phi = \mathcal{N}(\mathrm{Tr}_B\phi)\otimes \mathrm{Tr}_A\phi = \mathcal{N}(\rho)\otimes\rho \tag{11}$$

where $\rho = \frac{1}{d}I$. Also the entropy of each of the k^2 states $(\mathcal{N}\otimes\mathcal{I})(U_{j,k}\otimes\mathcal{I})\phi$ is $H((\mathcal{N}\otimes\mathcal{I})(\Phi_\rho))$, since each of the $(U_{j,k}\otimes\mathcal{I})(\phi)$ is a purification of ρ. Thus, we obtain the formula when $\rho = \frac{1}{d}I$.

The next step is to note that the inequality (9) also holds if the density matrix ρ is a projection onto any subspace of $\mathcal{H}_{\mathrm{in}}$. The proof is exactly the same as for $\rho = \frac{1}{d}I$. In fact, one can prove this case by restricting $\mathcal{H}_{\mathrm{in}}$ to the support of ρ, say $\mathcal{H}'$, and restricting $\mathcal{N}$ to act only on $\mathcal{H}'$.

For the next piece we need a little more notation. Assume we have a general ρ. Recall that we can assume that any quantum map $\mathcal{N}$ can be implemented via a unitary transformation $\mathcal{U}$ acting on the system $\mathcal{H}_{\mathrm{in}}$ and some environment $\mathcal{H}_{\mathrm{env}}$, where $\mathcal{H}_{\mathrm{env}}$ starts in some fixed initial state. We introduce $\mathcal{E}$, which is the completely positive map taking $\mathcal{H}_{\mathrm{in}}$ to $\mathcal{H}_{\mathrm{env}}$ by first applying $\mathcal{U}$ and tracing out everything but $\mathcal{H}_{\mathrm{env}}$. We then have

$$H(\mathcal{E}(\rho)) = H((\mathcal{N}\otimes\mathcal{I})\Phi_\rho)$$

for ρ a density matrix over $\mathcal{H}_{\mathrm{in}}$ and Φ_ρ is a purification of ρ.

As we use an argument involving typical subspaces; we first give some facts about typical subspaces. For technical reasons,[1] we will be using frequency-typical subspaces. For any ϵ and δ there is a large enough n such the Hilbert space $\mathcal{H}^{\otimes n}$ contains a typical subspace T (which is the span of typical eigenvectors of ρ) such that

[1]Our proof of Lemma 1 does not appear to work for entropy-typical subspaces unless these subspaces are modified by imposing a somewhat unnatural-looking extra condition.

1. $\operatorname{Tr} \Pi_T \, \rho^{\otimes n} \, \Pi_T > 1 - \epsilon$,

2. The eigenvalues λ of $\Pi_T \, \rho^{\otimes n} \, \Pi_T$ satisfy

$$2^{-n(H(\rho)+\delta)} \leq \lambda \leq 2^{-n(H(\rho)-\delta)} \ ,$$

3. $(1-\epsilon)2^{n(H(\rho)-\delta)} \leq \dim T \leq 2^{n(H(\rho)+\delta)}$

Let $T_n \subset \mathcal{H}^{\otimes n}$ be the typical subspace corresponding to $\rho^{\otimes n}$, and let ρ_{T_n} be the density matrix which is the projection onto T_n. It follows from well-known facts about typical subspaces that

$$\lim_{n\to\infty} \frac{1}{n} H(\rho_{T_n}) = H(\rho).$$

We can also show the following Lemma. We delay giving the proof of this lemma until after the proof of the theorem.

Lemma 1 *Let $\mathcal{N}$ be a noisy quantum channel and ρ a density matrix on the input space of this channel. Then we can find a sequence of frequency typical subspaces T_n corresponding to $\rho^{\otimes n}$, such that if ρ_{T_n} is the unit trace density matrix proportional to the projection onto T_n, then*

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{N}^{\otimes n}(\rho_{T_n})) = H(\mathcal{N}(\rho)). \tag{12}$$

Applying the lemma to the map onto the environment similarly gives

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{E}^{\otimes n}(\rho_{T_n})) = H(\mathcal{E}(\rho)). \tag{13}$$

Thus, if we consider the quantity

$$\frac{1}{n} \left[H(\rho_{T_n}) + H(\mathcal{N}^{\otimes n}(\rho_{T_n})) - H(\mathcal{E}^{\otimes n}(\rho_{T_n})) \right] \tag{14}$$

we see that it converges to

$$H(\rho) + H(\mathcal{N}(\rho)) - H(\mathcal{E}(\rho)), \tag{15}$$

as desired. This concludes the proof of the lower bound.

One more matter to be cleared up is the form of the prior entanglement to be shared by Alice and Bob. The most standard form of entanglement is maximally entangled pairs of qubits ("ebits"), and it is natural to use them as the entanglement resource in defining C_E. However, Eq. (8) involves the entangled state Φ_ρ, which typically not a product of ebits. This is no problem, because, as Lo and Popescu [18] showed, entangled pure states having equal entropy of entanglement can be interconverted not only with unit asymptotic efficiency, but in a way that requires an asymptotically negligible amount of (one-way) classical communication, compared to the amount of entanglement processed. Thus the definition of C_E is independent of the form of the entanglement resource, so long as it is a pure state. As it turns out, the lower bound proof below does not actually require construction of Φ_ρ itself, but merely a sequence of maximally entangled states on high-dimensional typical subspaces T_n of tensor powers of Φ_ρ. These maximally entangled states can be prepared from standard ebits with arbitrarily high fidelity and no classical communication [4].

2.2 Proof of Lemma 1

In this section, we prove

Lemma 1 *Suppose ρ is a density matrix over a Hilbert space $\mathcal{H}$ of dimension d, and $\mathcal{N}$, $\mathcal{E}$, are two trace-preserving completely positive maps. Then there is a sequence of frequency-typical subspaces $T_n \subset \mathcal{H}^{\otimes n}$ corresponding to $\rho^{\otimes n}$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \dim T_n = H(\rho), \tag{16}$$

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{N}^{\otimes n}(\pi_{T_n})) = H(\mathcal{N}(\rho)) \tag{17}$$

and

$$\lim_{n\to\infty} \frac{1}{n} H(\mathcal{E}^{\otimes n}(\pi_{T_n})) = H(\mathcal{E}(\rho)) \tag{18}$$

where π_{T_n} is the projection matrix onto T_n normalized to have trace 1.

For simplicity, we will prove this lemma with only the conditions (16) and (17). Altering the proof to also obtain the condition (18) is straightforward. We first need some notation. Let the eigenvalues and eigenvectors of ρ be λ_j and $|v_j\rangle$, with $1 \le j \le d$. Let the noisy channel $\mathcal{N}$ map a d-dimensional space to a d_{out}-dimensional space. Choose a Krauss representation for $\mathcal{N}$, so that

$$\mathcal{N}(\sigma) = \sum_{k=1}^{c} A_k \sigma A_k^\dagger\,,$$

where $c \le d^2$ and $\sum_{k=1}^{c} A_k^\dagger A_k = I$. Then we have

$$\mathcal{N}(\rho) = \sum_{j=1}^{d} \sum_{k=1}^{c} \lambda_j A_k |v_j\rangle\langle v_j| A_k^\dagger$$

We let

$$|u_{j,k}\rangle = \frac{1}{||A_k|v_j\rangle||} A_k|v_j\rangle \tag{19}$$

and

$$\mu_{j,k} = ||A_k|v_j\rangle||^2 \tag{20}$$

so that

$$\mathcal{N}(\rho) = \sum_{j=1}^{d}\sum_{k=1}^{c} \lambda_j \mu_{j,k} |u_{j,k}\rangle\langle u_{j,k}| \, . \tag{21}$$

We need notation for the eigenstates and eigenvalues of $\mathcal{N}(\rho)$. Let these be $|w_k\rangle$ and ω_k, $1 \leq k \leq d_{\text{out}}$. Finally, we define the probability p_{jk}, $1 \leq j \leq d$, $1 \leq k \leq d_{\text{out}}$, by

$$p_{jk} = \langle w_k| \mathcal{N}(|v_j\rangle\langle v_j|) \, |w_k\rangle \tag{22}$$

This is the probability that if the eigenstate $|v_j\rangle$ of ρ is sent through the channel $\mathcal{N}$ and measured in the eigenbasis of $\mathcal{N}(\rho)$, that the eigenstate $|w_k\rangle$ will be observed. Note that

$$\begin{aligned} \sum_j \lambda_j p_{jk} &= \langle w_k| \mathcal{N}\Big(\sum_j \lambda_j |v_j\rangle\langle v_j| \Big) \, |w_k\rangle \\ &= \langle w_k| \mathcal{N}(\rho) \, |w_k\rangle \\ &= \omega_k \end{aligned} \tag{23}$$

We now define the typical subspace $T_{n,\delta,\rho}$. Most previous papers on quantum information theory have dealt with entropy typical subspaces. We use frequency typical subspaces, which are quite similar, but have properties that make the proof of this lemma somewhat simpler.

A frequency typical subspace of $\mathcal{H}^{\otimes n}$ associated with the density matrix $\rho \in \mathcal{H}$ is defined as the subspace spanned by certain eigenstates of $\rho^{\otimes n}$. The eigenstates of $\rho^{\otimes n}$ are tensor product sequences of eigenvectors of ρ, that is, $|v_{\alpha_1}\rangle \otimes |v_{\alpha_2}\rangle \otimes \ldots \otimes |v_{\alpha_n}\rangle$. Let $|s\rangle$ be one of these eigenstates of $\rho^{\otimes n}$. We will say $|s\rangle$ is *frequency typical* if each eigenvector $|v_j\rangle$ appears in the sequence $|s\rangle$ approximately $n\lambda_j$ times. Specifically, an eigenstate $|s\rangle$ is *δ-typical* if

$$|\mathrm{N}_{|v_j\rangle}(|s\rangle) - \lambda_j n| < \delta n \tag{24}$$

for all j; here $\mathrm{N}_{|v_j\rangle}(|s\rangle)$ is the number of times that $|v_j\rangle$ appears in $|s\rangle$. The *frequency typical subspace* $T_{n,\delta,\rho}$ is the subspace of $\mathcal{H}^{\otimes n}$ that is spanned by all δ-typical eigenvectors $|s\rangle$ of $\rho^{\otimes n}$.

We define Π_T to be the projection onto the subspace T, and π_T to be this projection normalized to have trace 1, that is, $\pi_T = \frac{1}{\dim T}\Pi_T$.

¿From the theory of typical sequences [10], for any density matrix σ, any $\epsilon > 0$ and $\delta > 0$, one can choose n large enough so that

1. $\mathrm{Tr}\ \Pi_{T_{n,\delta,\sigma}}\, \sigma^{\otimes n}\, \Pi_{T_{n,\delta,\sigma}} > 1-\epsilon$.

2. The eigenvalues λ of $\Pi_{T_{n,\delta,\sigma}}\, \sigma^{\otimes n}\, \Pi_{T_{n,\delta,\sigma}}$ satisfy
$$2^{-n(H(\sigma)+\delta')} \leq \lambda \leq 2^{-n(H(\sigma)-\delta')}\ ,$$
where $\delta' = \delta d \log(\lambda_{\max}/\lambda_{\min})$, and $\lambda_{\max}$ ($\lambda_{\max}$) is the maximum (minimum) eigenvalue of σ.

3. $(1-\epsilon)2^{n(H(\sigma)-\delta')} \leq \dim T_{n,\delta,\sigma} \leq 2^{n(H(\sigma)+\delta')}$.

The property (1) follows from the law of large numbers, and (2), (3) are straightforward consequences of (1).

We first prove an upper bound that for all δ_1, and for sufficiently large n,
$$\frac{1}{n} H(\mathcal{N}^{\otimes n}(\pi_{T_{n,\delta_1,\rho}})) < H(\mathcal{N}(\rho)) + C\delta_1\,. \tag{25}$$
for some constant C. We will do this by showing that for any ϵ, there is an n sufficiently large such that we can take a typical subspace $T_{mn,\delta_2,\mathcal{N}(\rho)}$ in $\mathcal{H}_{\mathrm{out}}^{\otimes n}$ and project m signals from a source with density matrix $\mathcal{N}^{\otimes n}(\pi_{T_{n,\delta_1,\rho}})$ onto it, such that the projection has fidelity $1-\epsilon$ with the original source. Here, δ_2 (and δ_3, δ_4) will be linear functions of δ_1 (with the constant depending on σ, $\mathcal{N}$). By projecting the source on $T_{mn,\delta_2,\mathcal{N}(\rho)}$, we are performing Schumacher compression of the above source. ¿From the theorem on possible rates for Schumacher compression [16], this implies that
$$H(\mathcal{N}(T_{n,\delta_1,\rho})) \leq \lim_{m\to\infty} \frac{1}{m} \log \dim T_{nm,\delta_2,\mathcal{N}(\rho)}\,. \tag{26}$$
The property (3) above for typical subspaces then implies the result.

Consider the following process. Take a typical eigenstate
$$|s\rangle = |v_{\alpha_1}\rangle \otimes |v_{\alpha_2}\rangle \otimes \ldots \otimes |v_{\alpha_n}\rangle$$
of $T_{n,\delta,\rho}$. Now, apply a Krauss element A_k to each symbol $|v_{\alpha_j}\rangle$ of $|s\rangle$, with element A_k applied with probability $||A_k|v_{\alpha_j}\rangle||^2$. This takes
$$|s\rangle = \bigotimes_{j=1}^{n} |v_{\alpha_j}\rangle \tag{27}$$

to one of c^n possible states $|t\rangle$. Each state is associated with a probability of reaching it; in particular, the state

$$|t\rangle = \bigotimes_{j=1}^{n} |u_{\alpha_j,\beta_j}\rangle \tag{28}$$

is produced with probability

$$\tau = \prod_{j=1}^{n} \mu_{\alpha_j,\beta_j}. \tag{29}$$

Notice that, for any $|s\rangle$, if the $|t_z\rangle$ and τ_z are defined as in Eqs. (28) and (29), then

$$\mathcal{N}(|s\rangle\langle s|) = \sum_{z=1}^{c^n} \tau_z |t_z\rangle\langle t_z|. \tag{30}$$

where the sum is over all $|t\rangle$ in Eq. (28).

We will now see what happens when $|t_z\rangle$ when projected onto a typical subspace $T_{n,\delta_2,\mathcal{N}(\rho)}$ associated with $\mathcal{N}(\rho)^{\otimes n}$. We get that the fidelity of this projection is

$$\langle t_z|\Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}}|t_z\rangle = \sum_{|r\rangle\in T_{n,\delta_2,\mathcal{N}(\rho)}} \langle r|t_z\rangle\langle t_z|r\rangle, \tag{31}$$

where the sum is taken over all δ_2-typical eigenstates $|r\rangle$ of $\mathcal{N}(\rho)^{\otimes n}$. Now, we compute the average fidelity (using the probability distribution τ) over all states $|t_z\rangle$ produced from a given δ_1-typical eigenstate $|s\rangle$.

$$\begin{aligned}
\sum_z \tau_z \langle t_z|\Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}}|t_z\rangle &= \sum_{z=1}^{c^n} \sum_{|r\rangle\in T_{n,\delta_2,\mathcal{N}(\rho)}} \tau_z \langle r|t_z\rangle\langle t_z|r\rangle \\
&= \sum_{|r\rangle\in T_{n,\delta_2,\mathcal{N}(\rho)}} \langle r|\,\mathcal{N}(|s\rangle\langle s|)\,|r\rangle \\
&= \sum_{\substack{|r_z\rangle=\bigotimes_{j=1}^{n}|w_{\gamma_{z,j}}\rangle \\ |r_z\rangle\in T_{n,\delta_2,\mathcal{N}(\rho)}}} \prod_{j=1}^{n} p_{\alpha_j,\gamma_{z,j}} |r_z\rangle\langle r_z| \qquad (32)
\end{aligned}$$

The above quantity has a completely classical interpretation; it is the probability that if we start with the δ_1-typical sequence $s = \otimes|v_{\alpha_i}\rangle$, and take $|v_j\rangle$ to $|w_k\rangle$ with probability p_{jk}, that we end up with a δ_2-typical sequence of the $|w_k\rangle$.

We will now show that the projection onto $T_{n,\delta_2,\mathcal{N}(\rho)}$ of the average state $|t_z\rangle$ generated from a δ_1-typical eigenstate $|s\rangle$ of $\rho^{\otimes n}$ has expected trace at least $1-\epsilon$. This will be needed for the lower bound, and a similar result, using almost the same calculations, will be used for the upper bound. We know that the original sequence $|s\rangle$ is δ_1-typical, that is, each of the eigenvectors $|v_j\rangle$ appears approximately $n\lambda_j$ times. Now, the process of first applying the A_k to each of the symbols, and then projecting the result onto the eigenvectors of $\mathcal{N}(\rho)^{\otimes mn}$, takes $|v_j\rangle$ to $|w_k\rangle$ with probability p_{jk}. We start with a δ_1-typical sequence $|s\rangle$, so we have

$$\mathrm{N}_{|v_j\rangle}(|s\rangle) = (\lambda_j + \Delta_j)mn \tag{33}$$

where $|\Delta_j| < \delta_1$. Taking the state $|s\rangle = \bigotimes_j |v_j\rangle$ to $|r\rangle = \bigotimes_k |w_k\rangle$, and using Eq. (23), we get

$$\begin{aligned} \mathrm{E}\left(\mathrm{N}_{|w_k\rangle}(|r\rangle)\right) &= (\omega_k + \sum_j \Delta_j p_{jk})mn \\ &= (\omega_k + \Delta'_k)mn \end{aligned} \tag{34}$$

where $\Delta'_k \leq d\delta_1$. Now, by the law of large numbers, the quantities $\mathrm{N}_{|w_k\rangle}(|r\rangle)$ cannot be too far from their expectations. Let us take $\delta_2 = (d+1)\delta_1$. Then, by the law of large numbers, for every ϵ there are sufficiently large n so that $|r\rangle$ is δ_2-typical with probability $1-\epsilon$.

Now, we are ready to give the upper bound argument. We will be using the theorem about Schumacher compression [20, 16] that if, for all sufficiently large m, we can compress m states from a memoryless source emitting an ensemble of pure states with density matrix σ onto a Hilbert space of dimension mH, and recover them with fidelity $1-\epsilon$, then $H(\sigma) \leq H$.

We first need to specify a source with density matrix $\mathcal{N}(\pi_{T_{n,\delta_1,\rho}})$. Taking a random δ_1-typical eigenstate $|s\rangle$ of $\rho^{\otimes n}$ (chosen uniformly from all δ_1-typical eigenstates), and premultiplying each of the tensor factors $|v_{\alpha_j}\rangle$ by some A_k to obtain a vector $|t\rangle$ gives us the desired source with density matrix $\mathcal{N}(\pi_{T_{n,\delta_1,\rho}})$. We next project a sequence of m outputs from this source onto the typical subspace $T_{mn,\delta_2,\mathcal{N}(\rho)}$. Let us analyze this process. First, we will specify a sequence $|\bar{s}\rangle$ of m particular δ_1-typical eigenstates $|\bar{s}\rangle = |s_1\rangle|s_2\rangle\cdots|s_m\rangle$. Because each of the components $|s_i\rangle$ of this state $|\bar{s}\rangle$ is δ_1-typical, $|\bar{s}\rangle$ is a δ_1-typical eigenstate of $\rho^{\otimes mn}$. Consider the ensemble of states $|t\rangle$ generated from any particular δ_1-typical $|\bar{s}\rangle$ by applying the A_k matrices to $|\bar{s}\rangle$. It suffices to show that the average fidelity obtained when this ensemble is projected onto $T_{mn,\delta_2,\mathcal{N}(\rho)}$, has fidelity $1-\epsilon$, where ϵ goes to 0 as m goes to ∞. This will prove the theorem, as by averaging over all

δ_1-typical states $|\bar{s}\rangle$ we obtain a source with density matrix $\mathcal{N}^{\otimes n}(\pi_{T_{n,\delta_1,\rho}})$ whose projection has average fidelity $1-\epsilon$. This implies, via the theorems on Schumacher compression, that

$$H(\mathcal{N}(\pi_{T_{n,\delta_1,\rho}})) \leq \lim_{m\to\infty} \frac{1}{m} \dim T_{mn,\delta_2,\mathcal{N}(\rho)} \leq n(H(\mathcal{N}(\rho)) + \delta_3) \tag{35}$$

where $\delta_3 = \delta_2 d_{\text{out}} \log(\omega_{\max}/\omega_{\min})$; here $\omega_{\max}$ ($\omega_{\min}$) is the maximum (minimum) non-zero eigenvalue of $\mathcal{N}(\rho)$. If we let δ_1 go to 0 as n goes to ∞, we obtain the desired bound. To apply this argument, we need to make sure that the rate at which ϵ goes to 0 is independent of $|\bar{s}\rangle$; this is a straightforward consequence of the law of large numbers.

We need now only show that the projection of the states $|t\rangle$ generated from $|s_1\rangle \cdots |s_m\rangle$ onto the typical subspace $T_{mn,\delta_2,\mathcal{N}(\rho)}$ has trace at least $1-\epsilon$. We know that the original sequence $|\bar{s}\rangle$ is δ_1-typical, that is, each of the eigenvectors $|v_i\rangle$ appears approximately $mn\lambda_i$ times. Thus, the same argument using the law of large numbers that applied to Eq. (32) also holds here, and we have shown the upper bound for Lemma 1.

We now give the proof of the lower bound. We use the same notation and some of the same ideas and machinery as in our proof of the upper bound. Consider the distribution of $|t_z\rangle$ obtained by first picking a random typical eigenstate $|s\rangle$ of $\rho^{\otimes n}$, and applying a matrix A_k to each symbol of $|s\rangle$, with A_k applied to $|v_j\rangle$ with probability $|A_k|v_j\rangle|^2$. This gives an ensemble of quantum states $|t_z\rangle$ with associated probabilities τ_z such that

$$\mathcal{N}(\pi_{T_{n,\delta_1,\rho}}) = \sum_z \tau_z |t_z\rangle\langle t_z| \tag{36}$$

The idea for the lower bound is to choose randomly a set T of size $W = n(H(\rho)-\delta_4)$ from the vectors $|t_z\rangle$, according to the probability distribution τ_z. We will take $\delta_4 = C\delta_1$ for some constant C to be determined later. We will show that with high probability the selected set T of $|t_z\rangle$ vectors satisfy the criteria of Hausladen et al [12] for having a decoding observable that correctly identifies a state $|t_z\rangle$ selected at random with probability $1-\epsilon$. This means that these states can be used to send messages with rate $n(H(\rho)-\delta_4) - 2\epsilon$, showing that the density matrix of their equal mixture $\rho_T = \frac{1}{|T|}\sum_{z\in T} |t_z\rangle\langle t_z|$ has entropy at least $n(H(\rho)-\delta_4)-2\epsilon$. However, the weighted average of these density matrices ρ_T over all sets T is $\mathcal{N}(\pi_{T_{n,\delta_1,\rho}}) = \sum_z \tau_z |t_z\rangle\langle t_z|$, where each ρ_T is weighted according to its probability of appearing. By concavity of von Neumann entropy, $H(\mathcal{N}(\pi_{T_{n,\delta_1,\rho}})) \geq n(H(\rho)-\delta_1) - 2\epsilon$, and we are done.

The only thing left to do is to give the proof that with high probability a randomly chosen set of size W of the $|t_z\rangle$ obeys the criterion of Hausladen et al. The Hausladen et al. protocol for decoding [12] is first to project onto a subspace, for which we will use the typical subspace $T_{n,\delta_2,\mathcal{N}(\rho)}$, and then use the square root measurement on the projected vectors. The proof in [12] only gives the expected probability of error, but it is straightforward to modify to show that the probability of error $P_{E,i}$ in decoding the i'th vector, $|t_i\rangle$, is at most

$$P_{E,i} \leq 2(1 - S_{ii}) + \sum_{j \neq i} S_{ij} S_{ji}, \tag{37}$$

where $S_{ii} = \langle t_i | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_i \rangle$ and $S_{ij} = \langle t_i | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_j \rangle$. We have already shown that the expectation of $1 - S_{ii}$ is small, for $|t_i\rangle$ obtained from any typical eigenstate $|s\rangle$ of $\rho^{\otimes n}$.

We need to give an estimate for the second term of (37). Taking expectations over all the $|t_j\rangle$, $j \neq i$, we obtain, since all the $|t_z\rangle$ are chosen independently,

$$\mathrm{E}\Big(\sum_{j \neq i} S_{ij} S_{ji} \Big) = W \sum_z \tau_z \langle t_i | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_z \rangle \langle t_z | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_i \rangle, \tag{38}$$

where W is the number of random codewords $|t_z\rangle$ we choose randomly. We now consider a different probability distribution on the $|t_z\rangle$, which we call τ'_z. This distribution is obtained by first choosing an eigenstate $|s\rangle$ of $\rho^{\otimes n}$ with probability proportional to its eigenvalue (rather than choosing uniformly among δ-typical eigenstates of $\rho^{\otimes n}$), and then applying a Krauss element A_k to each of its symbols to obtain a word $|t\rangle$ (as before, A_k is applied to $|v_j\rangle$ with probability $|A_k|v_j\rangle|^2$). Observe that $\tau_z < 2^{2\delta' n} \tau'_z$, where $\delta' = d\delta \log(\lambda_{\max}/\lambda_{\min})$. This holds because the difference between the two distributions τ and τ' stems from the probability with which an eigenstate $|s\rangle$ of $\rho^{\otimes n}$ is chosen; from the properties of typical subspaces, the eigenvalue of every typical eigenstate $|s\rangle$ of $\rho^{\otimes n}$ is no more than $2^{-n(H(\rho)-\delta')}$, and the number of such eigenstates is at most $2^{n(H(\rho)+\delta')}$. Thus, we have

$$\begin{aligned}
\mathrm{E}\left(\sum_{j \neq i} S_{ij} S_{ji} \right) &= W \sum_z \tau_z \langle t_i | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_z \rangle \langle t_z | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_i \rangle \\
&\leq W 2^{2\delta' n} \sum_z \tau'_z \langle t_i | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_z \rangle \langle t_z | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_i \rangle \\
&= W 2^{2\delta' n} \langle t_i | \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} \mathcal{N}(\rho)^{\otimes n} \Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}} | t_i \rangle \\
&\leq W 2^{2\delta' n} 2^{-n(H(\rho)-\delta_3)}
\end{aligned} \tag{39}$$

where the last inequality follows from the bound (2) on the maximum eigenvalue of $\Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}}\mathcal{N}(\rho)^{\otimes n}\Pi_{T_{n,\delta_2,\mathcal{N}(\rho)}}$. Thus, if we make $W = 2^{n(H(\rho)-2\delta'-\delta_3-\delta)}$, we have the desired inequality (37), and the proof of Lemma 1 is complete.

2.3 Proof of the Upper Bound

We prove an upper bound of

$$C_E \leq \max_{\rho \in \mathcal{H}_{\text{in}}} \quad H(\rho) + H(\mathcal{N}(\rho)) - H(\mathcal{N} \otimes \mathcal{I}(\Phi_\rho)) \tag{40}$$

where Φ_ρ is a purification of ρ.

First, consider a channel $\mathcal{N}$ with entanglement-assisted capacity C_E and achieves capacity C_E. Then for every ϵ, there is a protocol that uses the channel $\mathcal{N}$ and some block length n, that achieves capacity $C_E - \epsilon$, and that does the following:

Alice and Bob start by sharing a pure entangled state ϕ, independent of the classical data Alice wishes to send. (Protocols where they start with a mixed state can easily be simulated by ones starting with a pure state, although possibly at the cost of additional entanglement.) Alice then performs some superoperator $\mathcal{A}_x$ on her half of ϕ to get $(\mathcal{A}_x \otimes \mathcal{I})(\phi)$, where $\mathcal{A}_x$ depends on the classical data x she wants to send. She then sends her half of $\mathcal{A}_x(\phi)$ through the channel $\mathcal{N}^{\otimes n}$ formed by the tensor product of n uses of the channel $\mathcal{N}$. Bob then possibly waits until he receives many of these states $(\mathcal{N}^{\otimes n} \otimes \mathcal{I})(\mathcal{A}_x \otimes \mathcal{I})(\phi)$, and applies some decoding procedure to them.

This is the definition of entanglement-assisted capacity, using only forward communication. If back-communication from Bob to Alice is not allowed, then Alice can do no better than encode all her classical information at once, by applying a single classically-chosen completely positive map $\mathcal{A}_x$ to her half of the entangled state ϕ, and then send it to Bob through the noisy channel $\mathcal{N}^{\otimes n}$. (If, on the contrary, back communication were allowed, it might be advantageous to use a protocol requiring several rounds of communication.) Note that the present formalism includes situations where Alice doesn't use ϕ at all, because the map $\mathcal{A}_x$ can completely discard all the information in ϕ.

In this section, we will assume that $\mathcal{A}_x$ is a unitary transformation $\mathcal{U}_x$. Once we have derived an upper bound assuming that Alice's transformations are unitary, we will use this upper bound to show that allowing her to use non-unitary transformations does not help her. This is proved by using the strong subadditivity property of von Neumann entropy; the proof will be deferred to the next section.

The next step is to apply the Holevo formula, Eq. (7), to the tensor product channel $\mathcal{N}^{\otimes n}$. For convenience, we let $\hat{\mathcal{N}} = \mathcal{N}^{\otimes n}$ denote the tensor product of many uses of the channel. In the x'th signal state, Alice sends her half of $(\mathcal{U}_x \otimes \mathcal{I})(\phi)$ through the channel, and Bob receives $(\hat{\mathcal{N}} \otimes \mathcal{I})(\mathcal{U}_x \otimes \mathcal{I})(\phi)$. Bob's state can be divided into two parts. The first of these parts is his half of ϕ, which, after Alice's part is traced out, is always in state $\mathrm{Tr}_A(\phi)$. The second part is the piece Alice sent through the channel, which, after Bob's part is traced out, is in state $\hat{\mathcal{N}}(\rho_x)$ where $\rho_x = \mathrm{Tr}_B(\mathcal{U}_x(\phi))$. Holevo's formula is applicable to this situation, as Bob is trying to decode information from a number of unentangled states. The first term of Holevo's formula is the entropy of the average signal, and this is bounded by

$$H(\hat{\mathcal{N}}(\sum_x p_x \rho_x)) + H(\rho_x). \tag{41}$$

The first term in this expression is the entropy of the average state that Bob receives through the channel, and the second term is the entropy of the state that Bob retained all the time. We use the subadditivity property of von Neumann entropy, that is, that the entropy of a joint system is bounded from above by the sum of the entropies of the two systems [19]. We can use $H(\rho_x)$ for the second term because Alice is using a unitary transformation to produce ρ_x from her half of the entangled state she shares with Bob, so the entropy $H(\rho_x)$ is the same for all x. Since we may assume that Alice and Bob share a pure quantum state, the entropy of Bob's piece is the same as the entropy of Alice's piece. Although it may not be the most obvious expression for this term, this expression will facilitate later manipulations of our formula.

The second term in Holevo's formula is the average entropy of the state Bob receives, and this is

$$\sum_x p_x H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_x})) \tag{42}$$

where Φ_{ρ_x} is a purification of ρ_x. This formula holds because after Alice's unitary transformation, Alice and Bob's entangled state is a purification of ρ_x, and any purification gives the same entropy of $H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_x}))$.

We thus get

$$n(C_E - \epsilon) \leq H\left(\hat{\mathcal{N}}(\sum_x p_x \rho_x)\right) + \sum_x p_x H(\rho_x) - \sum_x p_x H(\hat{\mathcal{N}} \otimes \mathcal{I}(\Phi_{\rho_x})) \tag{43}$$

However, by a lemma that is proved in the next section, the last two terms in this formula are a concave function of ρ_x, so we can move the sum inside these terms, and we get

$$C_E - \epsilon \leq \frac{1}{n} \Big(H(\hat{\mathcal{N}}(\rho)) + H(\rho) - H(\hat{\mathcal{N}} \otimes \mathcal{I}(\Phi_\rho)) \Big) \tag{44}$$

where

$$\rho = \sum_x p_x \rho_x .$$

Finally, the expression (8) for C_E is additive (this will be discussed in the next section), so that

$$C_E(\mathcal{N}_1 \otimes \mathcal{N}_2) = C_E(\mathcal{N}_1) + C_E(\mathcal{N}_2). \tag{45}$$

Using this, we can set $n = 1$ in Eq. (44), thus replacing $\hat{\mathcal{N}} = \mathcal{N}^{\otimes n}$ by $\mathcal{N}$. Since this equation holds for any $\epsilon > 0$, we obtain the desired formula (40).

2.4 Proofs of the Lemmas

This section discusses three lemmas needed for the previous section. The first of these shows that without loss of capacity, Alice can use a unitary transform for encoding. The next shows that the last two terms of the formula for C_E in Eq. (8) are a convex function of ρ. The last lemma shows that the formula for C_E is additive. Originally, we had fairly complicated proofs of all three of these lemmas, and we used the property of strong subadditivity for the first two. However, Prof. Holevo has pointed out that a much simpler proof than our original proof for the third lemma was already in the literature, and we will merely cite it.

For the proof of the first two lemmas in this section, we need the strong subadditivity property of von Neumann entropy [19]. This property says that if A, B, and C are quantum systems, then

$$H(\rho_{AB}) + H(\rho_{AC}) \geq H(\rho_{ABC}) + H(\rho_A). \tag{46}$$

It turns out to be a surprisingly strong property.

We need to show that if Alice uses non-unitary transformations $\mathcal{A}_x$, then she can never do better than the upper bound Eq. (40) we derived by assuming that she uses only unitary transformations $\mathcal{U}_x$. Recall that any non-unitary transformation $\mathcal{A}_x$ on a Hilbert space $\mathcal{H}_{\text{in}}$ can be performed by using a unitary transformation $\mathcal{U}_x$ acting on the Hilbert space $\mathcal{H}_{\text{in}}$ augmented by an ancilla $\mathcal{H}_{\text{anc}}$, and then tracing out the ancilla. We can assume that $\dim \mathcal{H}_{\text{anc}} = (\dim \mathcal{H}_{\text{in}})^2$.

What we will do is take the channel $\mathcal{N}$ we were given, that acts on a Hilbert space $\mathcal{H}_{\text{in}}$ and simulate it by a channel $\mathcal{N}'$ that acts on a Hilbert space $\mathcal{H}_{\rm in} \otimes \mathcal{H}_{\rm anc}$ where $\mathcal{N}'$ first traces out $\mathcal{H}_{\rm anc}$ and then applies $\mathcal{N}$ to the residual state on $\mathcal{H}_{\rm in}$. We can then perform any transformation S_x by performing a unitary operation $\mathcal{U}_x$ on $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm anc}$ and tracing out $\mathcal{H}_{\rm anc}$. Since we proved the formula Eq. (40) for unitary transformations in the previous section, we can calculate C_E by applying this formula to the channel $\mathcal{N}'$. What we show below is that the same formula applied to $\mathcal{N}$ gives a quantity at least as large.

Lemma 2 *Suppose that $\mathcal{N}$ and $\mathcal{N}'$ are related as described above. Let us define*

$$C = \max_{\rho\in\mathcal{H}_{\rm in}} H(\rho) + H(\mathcal{N}(\rho)) - H(\mathcal{N}\otimes\mathcal{I}(\Phi_\rho)) \tag{47}$$

and

$$C' = \max_{\rho'\in\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm anc}} H(\rho') + H(\mathcal{N}'(\rho')) - H(\mathcal{N}'\otimes\mathcal{I}(\Phi'_\rho)). \tag{48}$$

Then $C \geq C'$.

Proof: To avoid double subscripts in the following calculations, we now rename our Hilbert spaces as follows. Let $A = \mathcal{H}_{\rm in}$; $A' = \mathcal{H}_{\rm anc}$; $B = \mathcal{H}_{\rm out}$; and $E = \mathcal{H}_{\rm env}$. Let ρ' maximize C' in the above formula. We let $\rho = \mathrm{Tr}_{A'}\rho'$. Clearly, the middle terms in the above two formulae (47) and (48) are equal, since $\mathcal{N}(\rho) = \mathcal{N}'(\rho')$.

We now need to show that inequality holds for the first and last terms in C and C'; that is, we need to show

$$H(\rho) - H((\mathcal{N}\otimes\mathcal{I})(\Phi_\rho)) \geq H(\rho') - H((\mathcal{N}'\otimes\mathcal{I})(\Phi_{\rho'})) \tag{49}$$

Recall, we have a noisy channel $\mathcal{N}$ that acts on Hilbert space A, and the channel $\mathcal{N}'$ that acts on Hilbert space $A\otimes A'$ by tracing out A' and then sending the resulting state through $\mathcal{N}$. We also need to give purifications Φ_ρ and $\Phi_{\rho'}$ of ρ and ρ', respectively. Note that we can take $\Phi_\rho = \Phi_{\rho'}$, since any purification of ρ' is also a purification of ρ. Let us take these purifications over a reference system R. Consider the diagram in Figure 2. In this figure, $\rho_A = \rho$, $\rho_{AA'} = \rho'$ and $\rho_{AA'R} = |\Phi_\rho\rangle\langle\Phi_\rho| = |\Phi_{\rho'}\rangle\langle\Phi_{\rho'}|$. Then $\mathcal{N}$ maps the space A to the space B and $\mathcal{N}'$ maps the space AA' to the space B by tracing out A' and performing $\mathcal{N}$.

We have $H(\rho) = H(\rho_A) = H(\rho_{A'R})$, and $H(\rho') = H(\rho_{AA'}) = H(\rho_R)$. We also have $H((\mathcal{N}\otimes\mathcal{I})(\Phi_\rho)) = H(\rho_{A'RB})$ and $H((\mathcal{N}'\otimes\mathcal{I})(\Phi_{\rho'})) = H(\rho_{RB})$.

Thus,

$$\begin{aligned} C - C' &\geq H(\rho) - H((\mathcal{N}\otimes\mathcal{I})(\Phi_\rho)) - H(\rho') + H((\mathcal{N}'\otimes\mathcal{I})(\Phi_{\rho'})) \\ &= H(\rho_{A'R}) - H(\rho_{A'RB}) - H(\rho_R) + H(\rho_{RB}) \\ &\geq 0 \end{aligned} \tag{50}$$

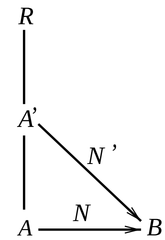

Figure 2: A is the input space for the original map $\mathcal{N}$. $A \cup A'$ is the input space for the map $\mathcal{N}'$. The output space for both maps is B. The space R is a reference system used to purify states in A and A'.

by strong subadditivity, and we have the desired inequality.

For the next lemma, we need to prove that the function

$$H(\rho) - H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_\rho))$$

is concave in ρ.

Lemma 3 *Let ρ_0 and ρ_1 be two density matrices, and let $\rho = p_0\rho_0 + p_1\rho_1$ be their weighted average. Then*

$$\begin{aligned} H(\rho) - H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_\rho)) \quad \geq \quad & p_0(H(\rho_0) - H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_0}))) \\ & + p_1(H(\rho_1) - H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_1}))) \end{aligned} \quad (51)$$

Proof: We again give a diagram; see Figure 3. Here we let the states be as follows: $\rho_A = \rho = p_0\rho_0 + p_1\rho_1$, so A is in the state ρ. We let R be a reference system with which we purify the states ρ_0 and ρ_1. Consider purifications $\Phi_0 = |\phi_0\rangle\langle\phi_0|$ and $\Phi_1 = |\phi_1\rangle\langle\phi_1|$ of ρ_0, ρ_1, respectively. Then we have

$$\rho_{AR} = p_0|\phi_0\rangle\langle\phi_0| + p_1|\phi_1\rangle\langle\phi_1| \quad (52)$$

We now let C_1 and C_2 be qubits which tell whether the system A is in state ρ_0 and ρ_1, and we will purify the system ρ_{AR} in the system ARC_1C_2 in the following way:

$$\phi_{ARC_1C_2} = \sqrt{p_0}|\phi_0\rangle|0\rangle|0\rangle + \sqrt{p_1}|\phi_1\rangle|1\rangle|1\rangle \quad (53)$$

Tracing out C_2, we get that the state of ARC_1 is

$$\rho_{ARC_1} = p_0|\phi_0\rangle\langle\phi_0| \otimes |0\rangle\langle 0| + p_1|\phi_1\rangle\langle\phi_1| \otimes |1\rangle\langle 1|, \quad (54)$$

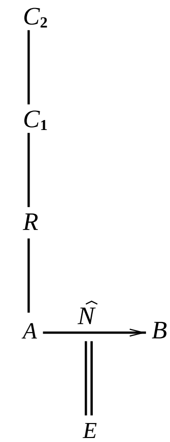

Figure 3: Here A is a Hilbert space we send through the channel $\hat{\mathcal{N}}$, and B is the output space. This mapping $\hat{\mathcal{N}}$ can be made unitary by adding an environment space E. We let R be a reference system which purifies the systems ρ_0 and ρ_1 in A, and C_1 and C_2 be two qubits purifying AR as described in the text.

so now C_1 can be thought of as a classical bit telling which of Φ_0 or Φ_1 is the state of the system AR. Note that we have the same expression after tracing out C_2.

Now, it's time for our analysis. We want to show equation (51) above. Notice that

$$H(\rho) = H(\rho_A) = H(\rho_{RC_1C_2})$$

and

$$H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_\rho)) = H(\rho_{BRC_1C_2}).$$

Now, suppose we have a classical bit C which tells whether a quantum system H is in state ρ_0 or ρ_1, with probability p_0 and p_1 respectively. The following formula for the expectation of the entropy of H is fairly easy to prove (this is analogous to the chain rule for the entropy of classical systems):

$$E(\rho_H) = p_0 H(\rho_0) + p_1 H(\rho_1) = H(\rho_{HC}) - H(\rho_C). \tag{55}$$

Using this formula, we see that

$$p_0 H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_0})) + p_1 H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_1})) = H(\rho_{BRC_1}) - H(\rho_{C_1}) \tag{56}$$

and

$$p_0 H(\rho_0) + p_1 H(\rho_1) = H(\rho_{AC_2}) - H(\rho_{C_2}) = H(\rho_{RC_1}) - H(\rho_{C_2}). \tag{57}$$

Putting everything together, we get

$$
\begin{aligned}
H(\rho) - H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_\rho)) &- \sum_{j=0}^{1} p_j \left(H(\rho_j) - H((\hat{\mathcal{N}} \otimes \mathcal{I})(\Phi_{\rho_j})) \right) \\
&= H(\rho_{RC_1C_2}) - H(\rho_{BRC_1C_2}) - H(\rho_{RC_1}) + H(\rho_{BRC_1}) \qquad (58)
\end{aligned}
$$

which is positive by strong subadditivity. Note that we used the equality $H(\rho_{C_1}) = H(\rho_{C_2})$, which holds by symmetry. This concludes the proof of Lemma 3.

The final lemma we need shows that we can set $n = 1$ and replace $\hat{\mathcal{N}} = \mathcal{N}^{\otimes n}$ by $\mathcal{N}$ in Eq. (44). This follows from the fact that C_E is additive, that is, if C_E is taken to be defined by Eq. (8), then

$$
C_E(\mathcal{N}_1 \otimes \mathcal{N}_2) = C_E(\mathcal{N}_1) + C_E(\mathcal{N}_2) \qquad (59)
$$

The $\geq$ direction is easy. We originally had a rather unwieldy proof for the $\leq$ direction based on explicitly expanding the formula for C_E and differentiating; However, A. Holevo has pointed out to us that a much simpler proof is given in [8], so we will spare the readers our proof.

3 Examples of C_E for Specific Channels

3.1 Gaussian Channels

The Gaussian channel is one of the most important continuous alphabet classical channels, and we briefly review it here. We describe the classical complex Gaussian channel, as this is the most analogous to the quantum Gaussian channel. For a detailed discussion of this channel see an information theory text such as [9, 10].

A classical complex Gaussian channel N of noise N is defined by the mapping in the complex plane

$$
\mathrm{N} : z \mapsto z' , \quad z' \sim G_N(z' - z) \ , \qquad (60)
$$

where the noise G_N is a Gaussian of mean 0 and variance N, i.e.,

$$
G_N(z) = \frac{1}{\pi N} e^{-|z|^2/N}. \qquad (61)
$$

Without any further conditions, the capacity of this channel would be unlimited, because we could choose an infinite subset of inputs arbitrarily far apart so that the corresponding outputs are distinguishable with arbitrarily small probability of error. The most natural constraint is that of average input signal power or energy, say S. That is, given an input distribution $W(x)$,

$$\int |z|^2 W(z)\, d^2z \leq S. \tag{62}$$

This complex Gaussian channel is equivalent to two parallel real Gaussian channels. It follows that the capacity of the complex Gaussian channel with average input energy S and noise N is

$$C_{\mathrm{Shan}} = \log\left(1 + \frac{S}{N}\right), \tag{63}$$

which is twice the capacity of a real Gaussian channel with average input energy S and noise N.

Before we proceed to discuss the quantum Gaussian channel, let us first review some basic results from quantum optics. In the quantum theory of light, each mode of the electromagnetic field is treated as a quantum harmonic oscillator whose commutation relations are the same as those of $SU(1,1)$. A detailed treatment of these concepts is available in the book [24]. The Hilbert space corresponding to a mode is countably infinite. A countable orthonormal basis for this space is the number basis of states $|n = j\rangle$, $j = 0, 1, 2, \ldots$, where the state $|n = j\rangle$ corresponds to j photons being present in the mode.

Another useful basis is that of the coherent states of light. Coherent states are defined for complex numbers α as

$$\begin{aligned} |\alpha\rangle &= D(\alpha)|0\rangle && (64)\\ &= e^{-|\alpha|^2/2} \sum_{j=0}^{\infty} \frac{\alpha^j}{\sqrt{j!}} |n = j\rangle && (65) \end{aligned}$$

where $D(\alpha)$ is the unitary displacement operator and $|0\rangle$ is the vacuum state containing no photons. The complex number α corresponds to the complex field vector of a mode in the classical theory of light. If $\alpha = x + ip$, then x is called the position coordinate and p the momentum coordinate. The displacement operator corresponds to displacing the complex number labeling the coherent state, along with a phase, i.e.,

$$D(\alpha)|\beta\rangle = |\alpha + \beta\rangle e^{i\,\mathrm{Im}(\alpha\beta^*)} \tag{66}$$

where Im takes the imaginary part of a complex number, i.e., $\mathrm{Im}(x+iy) = y$.

We also need thermal states, which are the equilibrium distribution of the harmonic oscillator for a fixed temperature. The thermal state with average energy S is the state

$$\begin{aligned} T_S &= \frac{1}{S+1}\sum_{j=0}^{\infty}\left(\frac{S}{S+1}\right)^j |n=j\rangle\langle n=j| \\ &= \frac{1}{\pi S}\int e^{-|z|^2/S}|z\rangle\langle z|\, d^2 z. \end{aligned} \tag{67}$$

The entropy of the thermal state T_S is

$$g(S) = (S+1)\log(S+1) - S\log(S)\,. \tag{68}$$

We are now ready to define the quantum analog of the classical Gaussian channel. (See [13] for a more detailed treatment of quantum Gaussian channels.) Since coherent states are an overcomplete basis, a quantum channel may be defined by its action on coherent states. We restrict our discussion to quantum Gaussian channels with one mode and no squeezing, which are those most analogous to classical Gaussian channels. These channels have an attenuation/amplification parameter k, and a noise parameter N. The channel amplifies the signal (necessarily introducing noise) if $k > 1$, and attenuates the signal if $k < 1$. The channel $\mathcal{N}$ with noise N and attenuation/amplification parameter k acts on coherent states as

$$\begin{aligned} \mathcal{N}(|\alpha\rangle\langle\alpha|) &= D_{k^2\alpha} T_N D^{\dagger}_{k^2\alpha} && \text{for } k \le 1 \\ \mathcal{N}(|\alpha\rangle\langle\alpha|) &= D_{k^2\alpha} T_{N+k^2-1} D^{\dagger}_{k^2\alpha} && \text{for } k \ge 1 \end{aligned} \tag{69}$$

The entanglement-assisted capacity of Gaussian channels was calculated in [13]. The density matrix ρ maximizing C_E is a thermal state of average energy S, and the entanglement-assisted capacity is given by

$$C_E = g(S) + g(S') - g(\frac{D+S'-S-1}{2}) - g(\frac{D-S'+S-1}{2}). \tag{70}$$

Here S is the average input energy; S' is the average output energy:

$$\begin{aligned} S' &= k^2 S + N && \text{for } k \le 1 \\ S' &= k^2 S + N + k^2 - 1 && \text{for } k \ge 1; \end{aligned} \tag{71}$$

and

$$D = \sqrt{(S+S'+1)^2 - 4k^2 S(S+1)}, \tag{72}$$

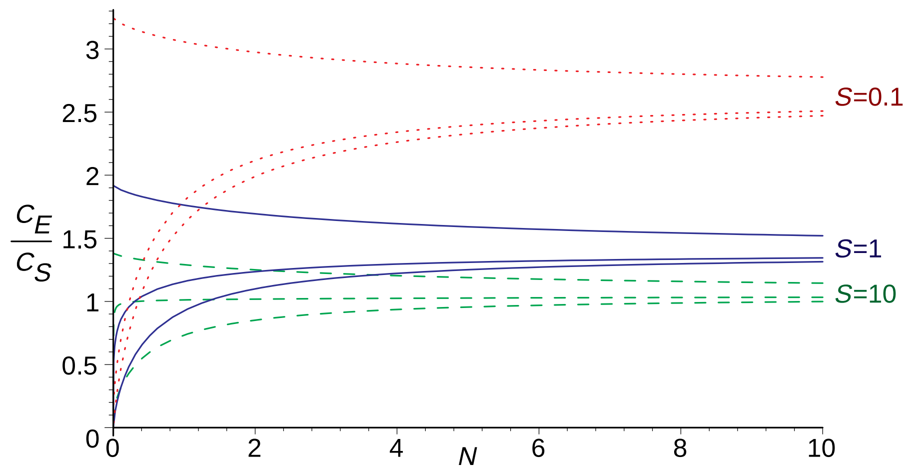

Figure 4: This figure shows the curves C_E/C_{Shan} for the nine combinations of values $k = 0.1$, 1, or 3; and $S = 0.1$, 1, or 10. The dotted curves have $S = 0.1$; the solid curves have $S = 1$; and the dashed curves have $S = 10$. Within each set, the curves have the values $k = 0.1$, $k = 1$, and $k = 3$ from bottom to top.

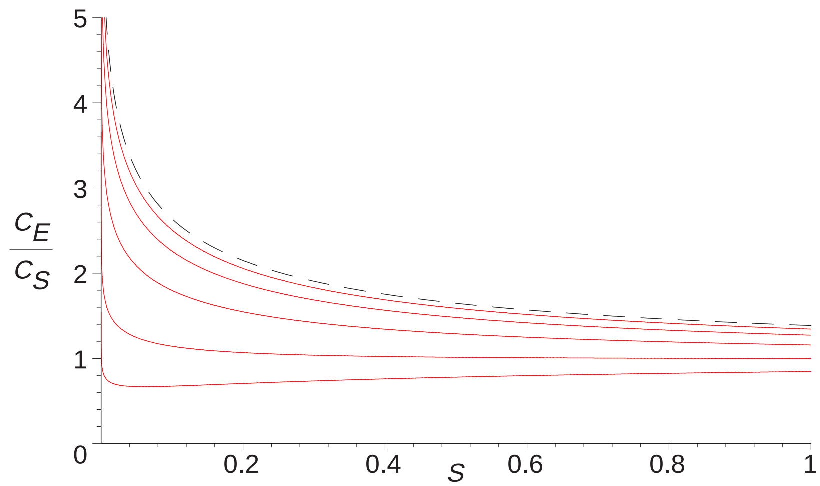

Figure 5: The solid curves show the ratio C_E/C_{Shan} for channels with $k = 1$ and $N = 0.1$, 0.3, 1, 3, and 10 (from bottom to top). The dashed curve is the limit of the solid curves as N goes to ∞; namely, $C_E/C_{\text{Shan}} = (S + 1)\log(1 + 1/S)$. These curves approach ∞ as S goes to 0, and approach 1 as S goes to ∞.

The first term of (70), $g(S)$, is the entropy of the input; the second term, $g(S')$, is the entropy of the output; and the remaining two terms of (70) are the entropy of a purification of the thermal state T_S after half of it has passed through the channel.

The asymptotics of this formula are interesting. Let us hold the signal strength S fixed, and let the noise N go to infinity. Then,

$$\lim_{N\to\infty} \frac{C_E}{C_{\mathrm{Shan}}} = (S+1)\log\left(1+\frac{1}{S}\right), \tag{73}$$

which is independent of the attenuation/amplification parameter k. This ratio shows that the entanglement-assisted capacity can exceed the Shannon formula by an arbitrarily large factor, albeit when the signal strength S is very small. We have plotted C_E/C_{Shan} for some parameters in Figs. 4 and 5.

Possibly a better comparison than that of C_E to C_{Shan} would be that of C_E to C_H, as C_H is the best rate known for sending classical information over a quantum channel without use of shared entanglement, However, the optimal set of signal states to maximize C_H for Gaussian channels is not known. For one-mode Gaussian channels with no squeezing, it is conjectured to be a thermal distribution of coherent states [13]; if this conjecture is correct, then $C_H \le C_{\mathrm{Shan}}$ for these channels, so the ratio C_E/C_{Shan} underestimates C_E/C_H; see Fig. 6.

Some simple bounds on C_E for the quantum Gaussian channel can be obtained using the techniques of [5]. Suppose that Alice takes a complex number α, encodes it as the state $|\alpha\rangle$, and sends this through a quantum Gaussian channel. Bob then measures it in the coherent state basis. Here, the measurement step adds 1 to the noise, and this channel is thus equivalent to a classical Gaussian channel with average received signal strength $k^2 S$, and average noise $N+1$ if $k \le 1$, $N+k^2$ if $k \ge 1$. The quantum Gaussian channel must then have capacity greater than the capacity of this classical Gaussian channel. Conversely, Alice and Bob can simulate a quantum Gaussian channel by using a classical complex Gaussian channel: Alice measures her state (in the coherent state basis), sends the result through the classical channel, and Bob prepares a coherent state that depends on the signal he receives. If Alice starts with a state $|\alpha\rangle$, when she measures it, she obtains a complex number $\alpha+\epsilon$ where ϵ is a Gaussian with mean 0 and variance 1. She can then multiply by k^2 to get $k^2\alpha + k^2\epsilon$. To simulate the quantum Gaussian channel, she must send this state through a classical channel with noise $N-k^2$ if $k \le 1$, and $N-1$ if $k \ge 1$. This classical channel must then have classical capacity greater than C_E for the quantum Gaussian channel

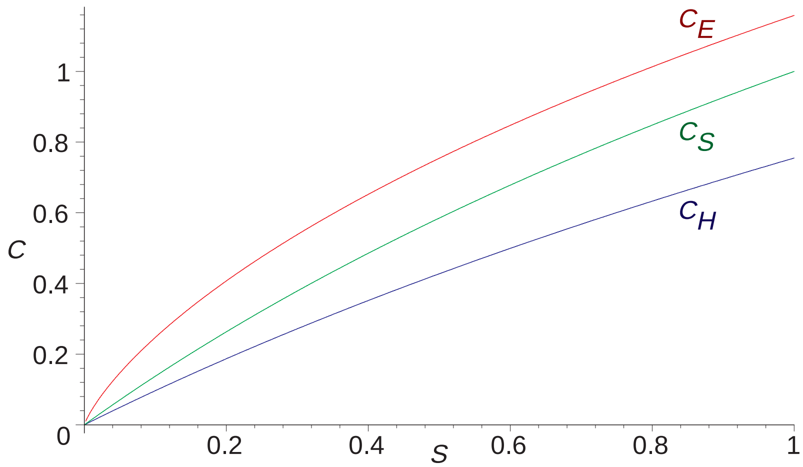

Figure 6: The values of C_E, C_{Shan}, and the conjectured C_H (in units of bits) are plotted for the Gaussian channel with noise $N = 1$ and no amplification or attenuation ($k = 1$). As the curves approach 0, their leading-order behavior is as follows: $C_H \approx S$, $C_{\text{Shan}} \approx (\log_2 e)S$, and $C_E \approx -\frac{1}{2} S \log_2 S$, so the ratios C_E/C_{Shan} and C_E/C_H approach ∞ as S goes to 0.

it is simulating. The arguments in this paragraph thus give bounds

$$\begin{aligned} \log\left(1 + \frac{k^2 S}{N+1}\right) \ &\leq \ C_E \ \leq \ \log\left(1 + \frac{S+1}{N/k^2 - 1}\right) \quad \text{for } k \leq 1 \\ \log\left(1 + \frac{S}{N/k^2+1}\right) \ &\leq \ C_E \ \leq \ \log\left(1 + \frac{k^2(S+1)}{N-1}\right) \quad \text{for } k \geq 1 \end{aligned} \tag{74}$$

If $k = 1$, we can compute better bounds than these based on continuous-variable quantum teleportation and superdense coding. Alice and Bob can use a shared entangled squeezed state to teleport a continuous quantum variable [6], and can also use such a state for a superdense coding protocol involving one channel use per shared state that increases the classical capacity of a quantum channel [7]. The squeezed state used, with squeezing parameter $r \geq 0$, is expressed in the number basis as

$$|s_r\rangle = \frac{1}{\cosh r} \sum_{j=0}^{\infty} (\tanh r)^j |n_A = j\rangle |n_B = j\rangle, \tag{75}$$

where n_A and n_B are the photon numbers in Alice's and Bob's modes, respectively. This state is squeezed, which means that it cannot be represented as a mixture of coherent states with positive coefficients. In this state, the uncertainty in the difference of Alice and Bob's position coordinates, $x_A - x_B$, is reduced, as is the uncertainty in the sum of their momentum coordinates $p_A + p_B$. The conjugate variables, $x_A + x_B$ and $p_A - p_B$, have increased uncertainty. If Alice and Bob measure their position coordinates, the difference of these coordinates is a Gaussian variable with mean 0 and variance e^{-2r}, while the sum is a Gaussian with mean 0 and variance e^{2r}. Similarly, if they measure their momentum coordinates, the sum has variance e^{-2r} while the difference has variance e^{2r}. Further, if either Alice's or Bob's state is considered separately, it is a thermal state with average energy $\sinh^2 r$.

In continuous-variable teleportation [6], Alice holds a state $|t\rangle$ she wishes to send to Bob, and one half of the shared state $|s_r\rangle$. She measures the difference of position coordinates of these states, $x_m = x_t - x_A$, and the sum of momentum coordinates, $p_m = p_t + p_A$. These are commuting observables, and so can be simultaneously determined. She sends these measurement outcomes to Bob, who then displaces his half of the shared state using $D(x_m + ip_m)$.

Using continuous-variable teleportation, Alice can simulate a quantum Gaussian channel with $k = 1$, average input energy S and noise N by sending the value $x_m + ip_m$ over a classical complex Gaussian channel with average input energy $S + (\cosh r)^2$ and noise $N - e^{-2r}$. This gives a bound equal to the classical capacity of this channel:

$$C_E \leq \log\left(1 + \frac{S + (\cosh r)^2}{N - e^{-2r}}\right). \tag{76}$$

Finding the r which minimizes this expression gives

$$e^{2r} = \frac{D_1 + 1}{N} \tag{77}$$

where

$$D_1 = \sqrt{(N+1)^2 + 4NS} \tag{78}$$

is the value of the variable D defined in Eq. (72) when we set $k = 1$. This gives the bound

$$C_E \leq \log\left(1 + \frac{S + (D_1 + N + 1)/(2N)}{N}\right). \tag{79}$$

Similarly, if Alice uses superdense coding [7] to send a continuous variable to Bob, her protocol simulates a classical Gaussian channel. The average energy input to this channel is $S - \sinh^2 r$ and the noise is $N + e^{-2r}$, so we obtain the bound

$$C_E \geq \log\left(1 + \frac{S - \sinh^2 r}{N + e^{-2r}}\right). \tag{80}$$

Maximizing this expression, we find the maximum is at $e^{2r} = (D_1 - 1)/N$, and the bound obtained is

$$C_E \geq \log\left(1 + \frac{S - (D_1 - N - 1)/(2N)}{N}\right). \tag{81}$$

Note that the bounds (76) and (80) reduce to the bounds of (74) when there is no entanglement in the squeezed state, i.e., when $r = 0$.

3.2 The Amplitude Damping Channel

The amplitude damping channel describes a channel which sends states which decay by attenuation from $|1\rangle$ to $|0\rangle$, but which do not undergo any other noise. This channel can be described by two Krauss operators,

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix} \quad \text{and} \quad \mathrm{A}_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$$

where

$$\mathcal{N} : \rho \rightarrow \sum_{j=1}^{2} A_j \rho A_j^\dagger \,.$$

The maximization over ρ to find C_E can be reduced to an optimization over one parameter, so it is numerically tractable, and the dependence of C_E and C_H on p are shown in Fig. 7. As the damping probability p goes to one, we can show that the Holevo capacity goes to

$$C_H \approx -\frac{1}{4}(1-p)\log(1-p) \tag{82}$$

and the entanglement-assisted capacity goes to

$$C_E \approx -(1-p)\log(1-p). \tag{83}$$

Here we use "$\approx$" to mean that the ratio of the two sides approaches 1 as p goes to 1. Thus, the ratio C_E/C_H approaches four. In our previous paper [5], we showed that for the qubit depolarizing channel, this ratio approached 3 as the depolarizing probability approached 1, and for the d-dimensional depolarizing channel, the ratio approached $d + 1$. We do not know whether this ratio is bounded for finite-dimensional channels, although we suspect it to be. If so, then an interesting question arises of how this bound depends on the dimensions $\dim \mathcal{H}_{\text{in}}$ and $\dim \mathcal{H}_{\text{out}}$.

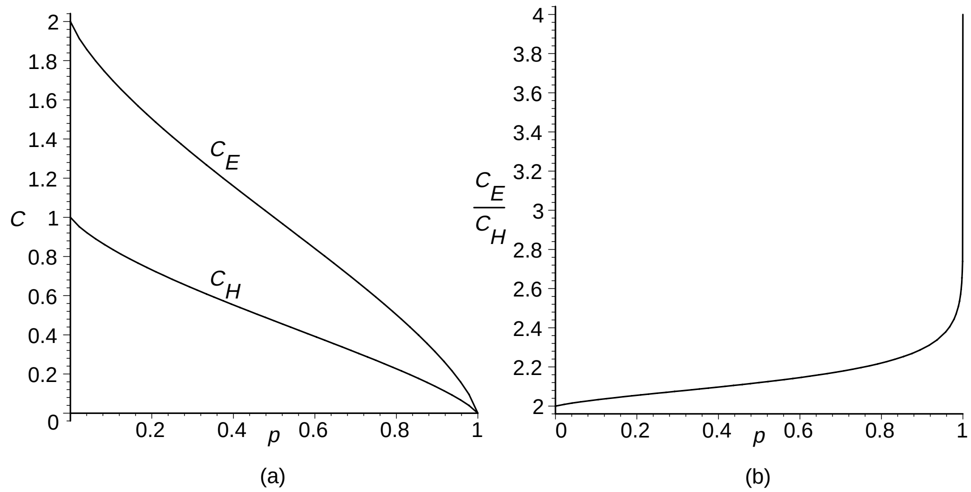

Figure 7: (a) The functions C_E and C_H for the amplitude damping channel are plotted against the damping probability p. (b) The ratio C_E/C_H is plotted. This curve is so steep near $p = 1$ that for $p = 1 - 10^{-50}$, the ratio C_E/C_H has reached only 3.8; the limiting value of 4 for $p = 1$ was derived analytically, and not numerically.

4 Classical Reverse Shannon Theorem

Shannon's celebrated noisy channel coding theorem established the ability of noisy channels to simulate noiseless ones, and allowed a noisy channel's capacity to be defined as the asymptotic efficiency of this simulation. The reverse problem, of using a noiseless channel to simulate a noisy one, has received far less attention, perhaps because noisy channels are not thought to be a useful resource in themselves (for the same reason, there has been little interest in the reverse technology of water desalination—efficiently making salty water from fresh water and salt). We show, perhaps unsurprisingly, that any noisy discrete memoryless channel of capacity C can be asymptotically simulated by C bits of noiseless forward communication from sender to receiver, given a source R of random information shared beforehand between sender and receiver. If this were not the case, characterization of the asymptotic properties of classical channels would require more than one parameter, because there would be cases where two channels of equal capacity could not simulate one another with unit asymptotic efficiency. In terms of the desalination analogy, water from two different oceans might produce equal yields of fresh drinking water, yet still not be equivalent because they produced unequal yields of partly saline water suitable, say, for car washing.

Although it is of some intrinsic interest as a result in classical information theory, we view the classical reverse Shannon theorem mainly as a heuristic aid in developing techniques that may eventually establish its quantum analog, namely the conjectured ability of all quantum channels of equal C_E to simulate one another with unit asymptotic efficiency.

Here we show that any classical discrete memoryless channel T, of capacity C, can be asymptotically simulated by C uses of a noiseless binary channel, together with a supply of prior random information R shared between sender and receiver.

The channel T is defined by its stochastic transition matrix T_{yx} between inputs $x \in \{1...d_I\}$ and outputs $y \in \{1...d_O\}$. Let T^n denote the extended channel consisting of n parallel applications of T, and mapping $x \in \{1...d_I^n\}$ to $y \in \{1...d_O^n\}$.

Theorem 2 (Classical Reverse Shannon Theorem) *For any classical discrete memoryless channel T, with Shannon capacity C, and for any block size n, there is a deterministic simulation protocol S_n for T^n which makes use of a noiseless forward classical channel and prior shared random information R, which for concreteness and without loss of generality may be taken to be a Bernoulli sequence (of course independent of the input x). The number of bits of forward communication on input $x \in \{1...d_I^n\}$ is a random variable; let its expectation (when R is chosen uniformly) be denoted $E(x,n)$. The simulation is exactly faithful in the sense that the stochastic matrix for S_n, when R is chosen randomly, is identical to that for T^n,*

$$\forall_{nxy} (S_n)_{yx} = (T^n)_{yx}, \tag{84}$$

and asymptotically efficient in the sense that

$$\lim_{n\to\infty} \frac{1}{n} \max_x E(x,n) = C. \tag{85}$$

To illustrate the central idea of the simulation, we prove the theorem first for a binary symmetric channel (BSC), then extend the proof to a general discrete memoryless channel. Let T be a binary symmetric channel of crossover probability p. Its capacity C is $1-H_2(p) = 1 + p\log_2 p + (1-p)\log_2(1-p)$. To prove the theorem in this case it suffices to show that for any $C' > C$, there is a sequence of simulation protocols S'_n, of expected communication cost $E'(x,n)$ such that

$$\forall_{nxy} (S'_n)_{yx} = (T^n)_{yx}, \tag{86}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \max_x E'(x,n) = C'. \tag{87}$$

The simulation protocol S'_n is as follows:

1. Before receiving the input $x \in \{0,1\}^n$, Alice and Bob use the random information R to choose a random set $Z(R,n)$ of $2^{nC'}$ n-bit strings.

2. Alice receives the n-bit input x.

3. Alice simulates the true channel T^n within her laboratory, obtaining an n-bit "provisional output" y. Although this y is distributed with the correct probability for the channel output, she tries to avoid transmitting y to Bob, because doing so would require n bits of forward communication, and she wishes to simulate the channel accurately while using less forward communication. Instead, where possible, she substitutes a member of the preagreed set $Z(R,n)$, as we shall now describe.

4. Alice computes the Hamming distance, $d = |x-y|$ between x and y.

5. Alice determines whether there are any strings in the preagreed set $Z(R,n)$ having the same Hamming distance d from x as y does. If so, she selects a random one of them, call it y', and sends Bob $0i$, where i is the approximately nC'-bit index of y' within the set $Z(R,n)$. If not, she sends Bob the string $1y$, the original unmodified n-bit string y, prefixed by a 1.

6. Bob emits y' or y, whichever he has received, as the final output of the simulation.

It can readily be seen that the probability of failure in step 5—i.e., of there being no string of the correct Hamming distance in the preagreed set $Z(R,n)$—decreases exponentially with n as long as $C' > C$. Thus the asymptotic classical communication cost approaches C' as required by Eq. (87). On the other hand, regardless of whether step 5 succeeds or fails, the final output is correctly distributed (satisfying Eq. (87)) since it has the correct distribution of Hamming distances from the input x, and for each Hamming distance, is equidistributed among all strings at that Hamming distance from x. The theorem follows.

For a general discrete memoryless channel the protocol must be modified to take account of the nonbinary input and output alphabets, and the fact that the output entropy may be different for different inputs, unlike the BSC case. The notion of Hamming distance also needs to be generalized. The new protocol uses the notion of *type class*[9, 10]. Two n-character strings belong to the same type class if they have equal letter frequencies (for example four a's, three b's, twelve c's etc.), and are therefore equivalent under some permutation of letter positions. We will consider input type classes (ITCs) and transition type classes (TTCs), the latter being defined as a set of input/output pairs (x,y) equivalent under some common permutation of the input and output letter positions. In other words, (x_1, y_1) and (x_2, y_2) belong to the same TTC if and only if there exists a permutation of letter positions, π, such that $\pi(x_1) = x_2$ and $\pi(y_1) = y_2$. Evidently, for any given input and output alphabet size, the number of ITCs, and the number of TTCs are each polynomial in n. Let $k = 1, 2...K_n$ index the ITCs, and $\ell = 1, 2...L_n$ the TTCs for inputs of length n. The TTC will be our generalization of the Hamming distance, since the transition probability $(T^n)_{yx}$ is equal for all pairs (x, y) in a given TTC. The new protocol follows:

1. Before receiving the input $x \in \{0, d_I - 1\}^n$, Alice and Bob use the common random information R to preagree on K_n random sets $\{Z(R, n, k) : k = 1...K_n\}$ of n-letter output strings. Each set has cardinality $2^{nC'}$, and there is a separate such set for each ITC k. In contrast to the BSC case, where the members of $Z(R, n)$ were chosen randomly from a uniform distribution on the output space, the elements of $Z(R, n, k)$ are chosen randomly from the (in general nonuniform) output string distribution induced by a uniform distribution of channel inputs over the k'th ITC.

2. Alice receives the n-letter input x, determines which ITC, k, it belongs to, and sends k to Bob, using $o(n)$ bits to do so.

3. Alice simulates the true channel T^n within her laboratory, obtaining an n-letter provisional output string y. Although this y is distributed with the correct probability for the channel input x, she tries to avoid transmitting y to Bob, because to do so would require too much forward communication. Instead she proceeds as described below.

4. Alice computes the index ℓ of the TTC to which the input/output pair (x, y) belongs. As noted above, this TTC index is the generalization of the Hamming distance, which we used in the BSC case.

5. Alice determines whether there are any output strings in the preagreed set $Z(R, n, k)$ having the same TTC index relative to x as y does. If

so, she selects a random one of them, call it y', and sends Bob the string $0i$ where i is the approximately nC'-bit index of y' within the set $Z(R, n, k)$. If not, she sends Bob the string $1y$.

6. Bob emits y' or y, whichever he has received, as the final output of the simulation.

This protocol deals with the problem of dependence of output entropy on input by encoding each ITC separately. Within any one ITC, the output entropy is independent of the input. The communication cost of telling Bob in which ITC the input lies is polylogarithmic in n, and so asymptotically negligible compared to n. Because one cannot increase the capacity of a channel by restricting its input, nC is an upper bound the maximal input:output mutual information achievable on any ITC. Moreover, for any ITC k and any input x in that ITC, the input:output pairs generated by the true channel T^n, will be narrowly concentrated, for large n, on TTCs whose transition frequencies approximate (to within $O(\sqrt{n})$ their asymptotic values. Therefore, as before, whenever $C' > C$, the probability of failure in step 5 will decrease exponentially with n. And as before, the simulated transition probability $(S_n)_{yx}$ on each ITC is exactly correct even for finite n. The reverse Shannon theorem for a general DMC follows.

5 Discussion—Quantum Reverse Shannon Conjecture

We conjecture (QRSC) that all quantum channels of equal C_E can simulate one another with unit asymptotic efficiency in the presence of unlimited entanglement between sender and receiver. By the results of the previous section, the conjecture holds for classical channels (where the shared random information required for the classical reverse Shannon theorem is obtained from shared entanglement). In our previous paper [5] we showed that the QRSC also holds for another class of channels, the so-called Bell-diagonal channels, which commute with teleportation and superdense coding. This can be seen by considering the two complementary scenarios in Figures 1b and 1c, which show respectively entanglement-assisted classical communication by a quantum channel, and entanglement-assisted classical simulation of a quantum channel. In Fig. 1b, Alice uses classical information x to modulate half of an entangled state ϕ, which is then sent through a quantum channel $\mathcal{N}$ and decoded by Bob to yield an effective noisy classical channel $\mathtt{N}$, whose capacity, we have argued, is $C_E(\mathcal{N})$, when $\phi, \mathcal{A}$ and $\mathcal{B}$ are chosen optimally. Conversely, in Fig. 1c Alice uses prior entanglement and a minimal amount of classical communication (thick line between Alice and Bob) to simulate the noisy channel $\mathcal{N}$. The amount of classical communication required is the entanglement assisted forward classical communication cost (FCCC) of the quantum channel $\mathcal{N}$. In [5] we showed that for Bell diagonal channels the channel's C_E is exactly equal to its entanglement-assisted FCCC. The QRSC asserts this equality holds, at least asymptotically, for all quantum channels. We hope that the arguments used to prove the classical reverse Shannon theorem can be extended to demonstrate this quantum analog of it.

One useful corollary to the quantum reverse Shannon conjecture would be a proof that a classical back channel from Bob to Alice cannot increase C_E. A causality argument shows that a back channel cannot increase C_E for noiseless quantum channels. If we could simulate noisy quantum channels by noiseless ones, this would imply that if a back channel increased C_E for any noisy channel, it would have to increase C_E for noiseless ones as well, violating causality.

We thank David DiVincenzo, Alexander Holevo, Michael Nielsen, and Barbara Terhal for helpful discussions.

References

[1] A. Ashikhmin and E. Knill, "Nonbinary quantum stabilizer codes," LANL eprint quant-ph/0005008.

[2] H. Barnum, M.A. Nielsen, and B. Schumacher, "Information transmission through noisy quantum channels," *Phys. Rev. A* **57,** pp. 4153–4175 (1998), LANL eprint quant-ph/9702049.

[3] H. Barnum, J. A. Smolin, and B. M. Terhal, "The quantum capacity is properly defined without encodings," *Phys. Rev. A* **58,** pp. 3496–3501 (1998) LANL e-print quant-ph/9711032.

[4] C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A* **53,** pp. 2046–2052 (1996).

[5] C.H. Bennett, P.W. Shor, J.A. Smolin, and A.V. Thapliyal, "Entanglement-assisted capacity of noisy quantum channels," *Phys. Rev. Lett.* **83**, pp. 3081–3084 (1999), LANL eprint quant-ph/9904023.

[6] S.L. Braunstein and H.J. Kimble, "Teleportation of continuous quantum variables," *Phys. Rev. Lett.* **80**, pp. 869–872 (1998).

[7] S. L. Braunstein and H. J. Kimble, "Dense coding with continuous quantum variables," *Phys. Rev. A* **61,** art. 042302 (2000).

[8] N. Cerf and G. Adami, "On the von Neumann capacity of noisy quantum channels, *Phys. Rev. A* **56,** pp. 3470–3483 (1997).

[9] Cover and Thomas, *Elements of Information Theory,* John Wiley and Sons (1991).

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems,* Akadémiai Kiadó, Budapest (1981).

[11] D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin and B. M. Terhal, "Unextendible product bases, uncompletable product bases, and bound entanglement," LANL eprint quant-ph 9908070.

[12] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A* **54,** pp. 1869–1876 (1996).

[13] A.S. Holevo and R.F. Werner "Evaluating capacities of bosonic Gaussian channels," *Phys. Rev. A* **63**, art. 032313, LANL eprint quant-ph/9912067.

[14] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Information Theory* **44,** pp. 269–273 (1998).

[15] P. Horodecki, M. Horodecki, and R. Horodecki, "Binding entanglement channels," *J. Modern Optics* **47,** pp. 347–354 (2000), LANL eprint quant-ph/9904092.

[16] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Modern Optics* **41**, pp. 2343–2350 (1994).

[17] G. Lindblad, "Quantum entropy and quantum measurements," in *Quantum aspects of optical communications, Lecture Notes in Physics 378*, C. Bendjaballah, O. Hirota, S. Reynaud (eds.) Springer, pp. 71–80 (1991).

[18] H.-K. Lo and S. Popescu, "The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource?" *Phys. Rev. Lett.* **83,** pp. 1459–1462 (1999), LANL eprint quant-ph/9902045.

[19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, UK, 2000.

[20] B. Schumacher, "Quantum coding," *Phys. Rev. A* **51**, pp. 2738–2747 (1995).

[21] B. W. Schumacher and M. A. Nielsen, “Quantum data processing and error correction”, *Phys. Rev. A* **54,** pp. 2629–2635, (1996).

[22] B. Schumacher and M. D. Westmoreland, ”Sending classical information via noisy quantum channels,” *Phys. Rev. A* **56**, pp. 131–138 (1997).

[23] G. Vidal and J. I. Cirac, “Irreversibility in asymptotic manipulations of entanglement,” LANL eprint quant-ph/0102036.

[24] D.F. Walls and G.J. Milburn, *Quantum Optics*, Springer Verlag, Berlin (1994).